The McGraw-Hill Companies

MEMS

About the Author

Dr. Nitaigour Premchand Mahalik received his B.Sc. (Engg.) in Electronics and Telecommunication Engineering in the year 1989 and Master of Engineering in Electronic Systems and Communication in December 1993 from the University College of Engineering (UCE), Burla. He did his Ph.D. in Mechatronics in June 1998 from the De Montfort University (DMU), U.K. and Postdoctoral research in June 2002 from Gwangju Institute of Science and Technology (GIST), South Korea. He joined in UCE, Burla functioning under Biju Patnaik University of Technology (BPUT), Orissa, as a lecturer in the Department of Electronics and Telecommunication Engineering in December 1990, then got promoted to the post of Senior Lecturer in 1994 and is presently continuing as Assistant Professor in the same Institute. He also worked as Controller of Examination at UCE, Burla.

Dr. Mahalik has academic interests in interdisciplinary engineering. The studies, teaching and researches in different departments and areas are the key to completing this interdisciplinary book. His teaching, research and development areas are mechatronics, robotics, automation, microinstrumentation, DCS and SCADA, micromanufacturing, industrial control and communications, augmented reality and simulation, rapid control prototyping and standardization.

He is the recipient of the prestigious National Overseas Scholarship from the Government of India for Ph.D. study abroad in the field of **Mechatronics** at DMU and Brain Korea – 21 Fellowship from the Government of South Korea for Postdoctoral research in the field of “Real-time Intelligent Distributed Machine Control Systems” in the Department of Mechatronics at GIST.

He was invited Visiting Professor to Moscow State Technological University, Russia, and GIST South Korea in the year 2001 and 2004. He has also visited University of Western Australia, Perth during 2003. Currently he is working as a Guest Professor in the department of Mechatronics, GIST, S. Korea.

Dr. Mahalik has completed writing and editing two books, “Mechatronics: Principles, Concepts and Applications” through McGraw-Hill, India and “Fieldbus Technology” through Springer-Verlag, Germany. Recently, he has completed editing a book “Micromanufacturing and Nanotechnology” through the Springer-Verlag, Germany. He is the author of more than 70 academic research papers that have already been published at the National and International levels.

He is a member and associated member of the various professional bodies such as ISTE (I), MMF (UK), AMIEE (UK), MMF (I), MISA (USA). He is the reviewer of the International Journal of American Society of Nuclear Energy, USA; International Association for Science and Technology Development, Canada; International Journal of Thermal Sciences, Netherland; International Journal of Computational Intelligence, Turkey; International Journal of Computer Applications in technology, Germany; and IEEE journal of Evolutionary Computation, USA. Dr. Mahalik is the Editors-in-Chief of the International Journal of Automation and Control (Inderscience), International Journal of Engineering Studies (RIP) and Guest editor of International Journal of Computer Applications in Technology and International Journal of Technology Management. Dr. Mahalik is also involved with the National Research Foundation, South Africa in their evaluation process.

The McGraw·Hill Companies

MEMS

Nitaigour Premchand Mahalik
Department of Industrial Technology
California State University
California, USA

Tata McGraw-Hill Publishing Company Limited
NEW DELHI

McGraw-Hill Offices
New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

The McGraw·Hill Companies

Tata McGraw-Hill

Published by the Tata McGraw-Hill Publishing Company Limited,
7 West Patel Nagar, New Delhi 110 008.

ISBN-13: 978-0-07-063445-9
ISBN-10: 0-07-063445-9

Head—Higher Education & School: *S. Raghothaman*
Executive Publisher-: *Vibha Mahajan*
Jr. Sponsoring Editor-: *Shukti Mukherjee*
Editorial Services Manager-: *Mini Narayanan*
Deputy General Manager—Higher Education & Sales: *Michael J. Cruz*
Asst. Product Manager-: *Biju Ganesan*
Controller—Production-: *Rajender P Ghansela*
Asst General Manager—Production: *B.L. Dogra*
Senior Production Executive: *Anjali Razdan*

Typeset at Tej Composers, WZ-391, Madipur, New Delhi 110063, and printed at Krishna Offset 618, B-3, Vishwas Nagar, Delhi 110032
Cover Printer: Meenakshi Printer
RXZYYRLYDXRAD

The McGraw·Hill Companies

The McGraw-Hill Companies

Dedicated to my wife

Ellyy

Contents

Preface

The technology that considers micromanufacturing of microscale transducers, actuators, probes, capacitors, inductors, valves, gears, pumps, gyroscope, mirrors, switches, and so on, similar to semiconductor chips is referred to as Microelectromechanical Systems, or MEMS in short. MEMS, an advanced product and equipment design concept, have already emerged in order to cater the development of miniaturized products. It has become the preferred choice for the development of next generation products and equipments, which are to be used for many kinds of applications. In essence, MEMS are small and integrated devices, which combine electronics, electrical as well as mechanical elements to meet the control related functional requirements. MEMS technology is an extended form of traditional microelectronic IC fabrication technique and an advanced process as far as micromanufacturing of microsystems is concerned. MEMS technology can fabricate capacitors and inductors as well as mechanical elements such as springs, gears, beams, diaphragms, and so on.

MEMS is a kind of Multiphysics-Multiengineering discipline and its scope is enormous in magnitude. Ideally MEMS combines Physics, Microelectronics, Micromechanics, Material Science, and Computer Aided Design (CAD) technology. Multiphysics-Multiengineering integration methods look at systems design that adheres the basic principles of microscale and nanoscale integration methods considering the physics, chemistry, biology and several engineering backbones. The technology also embeds computer aided simulation methods for the optimization of design parameters at all length scales, from the nanoscale to microscale to macroscale.

The subject MEMS is slowly being introduced for the engineering students at undergraduate (UG) and postgraduate (PG) levels all over the world. In many foreign Universities most of the engineering disciplines have included MEMS as a prime subject in their curriculum. The book can become a textbook at the UG and PG levels in most of the engineering disciplines including electronics, mechanical, instrumentation, industrial engineering, automation engineering, manufacturing, chemical engineering, material science and computer science and engineering. Further, a wide range of researchers will be able to find many useful tips and hints from this book. The way the subject "Mechatronics" presently, has found a place in the curriculum of all engineering disciplines, in future the subject MEMS will be included like wise. It is erotically appealing for engineering students to have greater ideas on this new subject. It will be effectual if all engineering students within the engineering school are given the scope of learning this weighty subject.

The amalgamation of various theories, principles, phenomenon, techniques and methodologies to cater the pressing needs have long been emerging as new disciplines. Fundamentals to the state-of-the-art field devices (sensors and actuators are called field devices), subsystems and modules must be understood along with the overview of underlying architectures and design approaches to defining MEMS. Specialized opportunity at the UG and PG level should embody source of concepts and techniques, which have recently been applied in practical situation. It is true that knowledge can be

acquired passively. But, students have to ripen their academic skills by actively learning from the valuable lessons. The book satisfies the requirement.

The technological education and research scenario, all over the world, is converging towards multi-discipline one. Worldwide technical education curriculum has been a reverse structure contrast to traditional methodology. Thus the present scenario is different as compared to recent past in the sense that the engineering disciplines are now dilating. The primary reason is being the fact that the current *technological designs* is of highly complex and inter-interdisciplinary nature involving *synergistic* integration of many aspects of engineering knowledge base. Since this book covers aspects of interdisciplinary subjects synergistically, the importance of the book is considered significant.

The importance of MEMS study, its emergence, definition and principles has been presented in first place (Chapter 1). In this part of the book, broad knowledge accommodating the principle of MEMS has been described sequentially, so that the readers can immediately grasp the topic at their fight sights. The scope of MEMS is also briefly discussed. Comprehensive in scope, and gentle in approach, the book will help the readers to achieve a thorough grasp of the basics and will enlighten the implementation concepts gradually.

Chapter two includes the fabrication processes and methods those are required for the MEMS design. Besides basic study in terms of definition the insight into the usefulness and applicability of such processes, methods and techniques are presented. In particular, the significance of lithography, surface micromachining, bulk micromachining, etching, thin film deposition, doping, wafer bonding, etc. is elucidated. This includes the study of UV lithography, X-ray lithography, E-beam lithography, soft lithography, LPCVD, PECVD, sputtering, diffusion doping, ion implantation, dry etching, wet etching, RIE, anodic bonding, fusion bonding, real estate gain, etc.

Chapter 3 deals with system modeling. The behaviour of the engineering system is studied through modeling and analysis. The important phases of modeling always starts with the identification and description of the physical systems and then formulation of a mathematical model. The systems are identified and described utilizing '*basic modeling elements*'. The basic modeling elements facilitate in the framing balancing model equation keeping in view the law of conservation of energy. Once the physical system is fully described, it becomes easier to study, analyze, evaluate, and predict its behavior and performance so that design requirements can be met. In the sequence, role of modeling, definitions, principles and methods are presented lucidly. The resulting models shape the foundation for understanding, studying and manipulating the behaviour of the systems. This chapter also presents the definitions of property parameters of materials.

From the viewpoint of integrated system, physically, a full-fledged MEMS device is composed of two main groups of functional subsystems; active and passive systems. Sensors and actuators are considered as active systems. Chapter 4 deals with the two types of passive systems: Passive Electronic Systems (PES) and Passive Mechanical Systems (PMS). Within MEMS, many signals require some amount of signal conditioning at the intermittent stage. Some of the signal conditioning functions are rectification, amplification, filtering, Analog to Digital Conversion (ADC), Digital to Analog Conversion (DAC), Isolation, multiplexing to name a few. In this respect, PES such as amplifiers, filters, data converters, isolators, clippers, bridge circuits, etc. are described. Similarly, mechanisms play very important role in designing MEMS. MEMS related mechanisms are called PMS which are gear, bearing, flexure, hinges, anchor, crank and slider, rack and pinion and so on.

Next, different types of fundamental components used for MEMS application are described. Chapter 5 provides an overview of mechanical transducers and actuators. These include basic mechanical

sensors and actuators, their principle of operations. Latest developments with regard to sensors and their application domain are narrated. In particular the chapter describes microplates, diaphragm structure, piezoelectric effects, piezomechanics, capacitive sensing and actuation, strain measurement, pressure measurement, application of microphone, MEMS gyroscope, inchworm technology based actuation, etc.

Chapter 6 is dedicated to describe the thermal MEMS. The chapter considers sensors and actuators design based on thermal effect. Thermal energy basics and heat transfer processes, various equipments like heat pump, pyroelectricity thermal flow sensors and actuators, diode and transistor temperature sensors are elaborately discussed. Their principles of operations and application domains are presented. Thermocouple, RTD, MEMS thermal probe, microhotplate, thermo vessels, SMA actuators, U-shaped horizontal and vertical actuator, MEMS relays, data storage cantilever etc. are the matter of studies.

MOEMS is a miniaturised system combining optics, micromechanics and microelectronics. MOEMS technology requires a different set of rules for operation when compared with the normal MEMS devices. Mostly, MOEMS have emerged to provide unparalleled functionality in telecommunication applications. Although the ultimate speed of these devices is unlikely to compete with solid-state electro-optic devices, the precision that can be achieved with MOEMS contributes to good performance and negligible signal degradation in the channels, thereby enabling a flexible all-optical system. System bandwidth and power consumption are the key issues. Using fiber links and optical methods, MOEMS technology demands precision interfaces and integrated components in order to achieve reliable and available quality of service (QoS). Chapter 7 deals with MOEMS devices which includes properties of light, beam splitter, microlens, micromirrors, Digital Light Processing and DMD devices, Silicon Light Machines and GLV technology, optical switch, waveguide and tuning, stress measurement, etc.

Chapter-7 discusses the magnetic microsystems, called MagMEMS. The magnetic sensor and actuator materials, the principle of Hall-effect, different magnetic components like resistor, diode and transistor are presented. The pressure sensor using MOKE, MagMEMS actuators for optical switching, bidirectional actuator principle and fabrication, large force reluctant actuator, magnetic probe based storage device, etc. are covered in this chapter.

Radio frequency microelectromechanical systems (RF MEMS) is a field that concerns with development of micromachined devices such as filters, oscillators, resonators, and switches, aimed at high frequency (~1Mhz to 60GHz) communication applications. Chapter 9 covers introduction to communication system, application areas of RF MEMS, MEMS inductor, capacitor, varactors (including fabrication process), tuners, patch resonator, microdisk resonator, bulkmode resonator, contact type and capacitive coupled MEMS switches, distributed phase shifters and their application, etc.

In Chapter 10 microfluidics is presented. The study of motion or transportation of fluids and their mixtures at a microscale level is known as microfluidics. Microdevices, which are used to transport and store fluid are called microfluidic systems (MFS) or microfluidic devices (MFD). Typically, the MFS handle fluid volumes on the order of nanoliter. Microfluidic systems are the subject of great scientific and commercial interest for a wide range of applications, including the biomedical, environmental, automotive, aerospace, and defense. The development of fluidic microsystems requires fundamental study on actuation mechanisms. Some of the important mechanisms are, Dielectrophoresis, Electrowetting, Electrothermal, Thermocapillary, Electroosmosis and optoelectrowetting. The theory and principle of operations are explained in proper order.

The scope of MEMS technology has recently been extended to health sciences and chemical industry. MEMS devices used for health science and chemical industry are called BioMEMS and chemical microsystems. Chapter-10 includes biomedical and chemical microsystems. The system classification and chemicals used in bioMEMS, different devices like DNA sensors and their application are illustrated. The sensing principle, type of active transducer material, chemoresistor, chemocapacitor, chemodiodes, ISFET, electronic nose, DNA chip, mass spectroscopy, fluorescence detection, CalSpec, SAW sensor, etc. are studied in this chapter.

The objective of this book is to provide detail information as far as proof-of-principle, concepts, design, development and applications of MEMS are concerned. NEMS (Nanoelectromechnical Systems) will be presented in one chapter (Chapter 12). It does this since MEMS has already emerged as a full-fledged technology whereas development of NEMS is still at its rudimentary stage. NEMS groups of products and systems fall under a well-known technology called *nanotechnology*. Nanotechnology is a technology that considers designing products at the nanoscale. This chapter is intended to present fundamental knowledge with respect to recent developments in nanotechnology. The properties of fullerene and various forms of CNTs are presented at the beginning. SWCNT, MWCNT, chiral vector, chiral angle, structure of CNTs, development of CNTs (laser ablation, electric arc discharge and CVD), applications of CNTs (quantum wire, transistors, X-ray materials for biomedical applications, artificial muscle (bucky paper) fuel cell, PEM, molecular machine components, etc.) are covered in this chapter.

Chapter 13 introduces simulation based design methodology and the concept of multiscale design approach. Simulation is a design methodology, using which the designers study the performance of the system (a MEMS structure, component, for example) prior to its real design. Based on fundamental governing laws the simulation study is carried out with the help of a computer. In particular, the need for simulation, FEM based simulation, design flow within the simulation environment, some important simulation tools such as HFSS, DS/MEMS and CA/MEMS, FEMPRO, ANSYS Multiphysics and SUGAR are presented. Multiscale design methodology such as MAAD, QCM, BDM, CGMD, and CADD are introduced. The chapter concludes with an illustration of multiscale system integration and the complexity involved in multiscale design approach.

Without discussion of the performance of MEMS devices, the book and the knowledge of the reader will remain unfulfilled. Chapter-14 is dedicated to understand the performance of the microdevices. Calibration technique and reliability of microdevices are also covered. This last chapter defines the performance measuring parameters in detail. In particular, repeatability, resolution, sensitivity, accuracy and precision, nonlinearity, transient property and specification, hysteresis, stability, reliability, availability, material characterization, etc. is made clear. The final part of the chapter discusses about the requirement of the array-based devices.

The book is dedicated for use at various levels (Diploma, UG, PG and even the Ph.D. research) in various departments including Instrumentation Engineering, Production Engineering, Mechanical Engineering, Manufacturing Engineering, Mechatronics Engineering, Electrical Engineering, Electronics Engineering, Communication Engineering and of course Interdisciplinary Engineering. This textbook is written for the beginners who wish to study and practice the subject of MEMS. The student will find it easy to understand and the teachers will get all the information on this subject at one place.

NITAIGOUR PREMCHAND MAHALIK

Acknowledgements

I appreciate the valuable comments given by the reviewers, namely Dr. M.Y. Khire of Walchand College of Engineering Sangli. They found time in their busy schedules to read the text and offer valuable suggestions for improvement.

I must express my profound sense of gratitude to my brothers Er. Babuchand Mahalik and Er. Kalachand Mahalik who had really planted the seed of knowledge in my brain.

I sincerely extend my thankfulness to University College of Engineering, Burla (functioning under Biju Patnaik University of Technology, Orissa, India) and Gwangju Institute of Science and Technology, South Korea for providing me all facilities for writing this book. My special thanks goes to all the valuable authors, researchers, companies, research laboratories, editors, publishers, etc. of the papers, books, articles those I have referred (enlisted towards the end of the book). Without their valuable work and research results this book could not have been completed. Every precaution has been taken to credit their work in terms of citing their names and copyright notices under the figures and texts incorporated in this book. I would like to acknowledge Axsun Technologies Inc., ACM SIGMETRICS, Addison-Wesley, American Chemical Society, bhv PUBLISHING HOUSE, Clarendon Press, Elsevier Science, Foresight Institute, Fujitsu, Horizon House Publications, IBM, IEE (UK), IEEE (USA), Institute of Physics, Intel, IOP Publishing, John Wiley & Sons, Kluwer Academic Publishers, Longman, Motorola Inc., NASA, Nature, OSA, Honeywell, Prentice Hall, Royal Society of Chemistry, Russian Academy of Sciences, SEIKO, SPIE, Sony Inc., Springer-Verlag, Wiley-Interscience and others whose names have been mentioned in the reference list.

I render my thankfulness to Sri Abhimanyu Mahalik (my father), Sri Khageswara Mohanty, Col. K. P. Singh Deo, Sri Rudra Pani, Sri Tathagata Satapathy, Sri Viswa Bhusana Harichandana, Smt. Minu Mahapatra, Indramani Behera and many others for their valuable support and help (directly or indirectly) to complete this book, while staying in GIST, South Korea.

I also owe the deep sense of love to my students from whom I have gathered so much of feedback during my teaching and interactions.

Last but not the least I want to thank my two sons Likun and Pikun for sacrificing their valuable moments and making it all worthwhile.

Nitaigour Premchand Mahalik

Chapter 1

Introduction

Objectives

The objective of this chapter is to introduce the following topics which will be elaborately covered in detail in the rest of the book.

- The origin of Microelectromechanical Systems (MEMS)
- Manufacturing and micromanufacturing
- Micromilling methods and exemplar applications
- Introduction to MEMS and System-on-a-Chip (SOC)
- The density related information of MEMS system vis-à-vis application areas
- The market growth of MEMS technology
- Introduction to microfabrication and micromachining
- Difference between bulk and surface micromachining
- Microsensors and microactuators
- Transduction principles
- Introduction to mechanical, thermal, magnetic and electrostatic actuators
- Radio Frequency (RF) MEMS and Micro Opto Electro Mechanical Systems (MOEMS)
- Bio, Chemo and Microfluidic MEMS
- The scope of the book

1.1 BACKGROUND AND INTRODUCTION

In the year 1959, Feynman proposed the possibility of manufacturing ultraminiaturize systems for a variety of applications, to a level that may involve multiscale formulation methods in terms of manipulation of molecules and atoms. Although the idea was visionary at the time, the world has moved in the apposite direction. Sophisticated miniaturized components and systems may indeed change products and equipment in the most dramatic way. Methodology and design prospects of miniaturized products and systems represent a broad topical subject, including fundamental physics, martial science, computing methods, ultra-precision engineering, fabrication technology and micromachining based upon the principles, characterization, modeling, simulation and state-of-the-art technology.

MEMS (Microelectromechanical Systems), an advanced product and equipment design concept, has already emerged in order to cater to the development of miniaturized products (Refer to Appendix E to see some of the MEMS basic products presented schematically). It has becoming the preferred scenario

for the next generation sophisticated products and equipment which are to be used for many kinds of applications. Research and development (R&D) in the field of MEMS has been underway since a decade. Technological researches are now dilating. Many industries, in conjunction with academic institutions and R&D sectors, have been investing good amount of money and resources in pursuing technological growths in this field. As the design phases of MEMS is highly complex, multiphysics and inter-interdisciplinary in nature, entailing synergistic integration of many aspects of fundamental knowledge base, the pursuance of unified understanding on related sub-domains is of paramount importance.

With respect to the past and the present, arguably, it can be stated that the R&D activities on miniaturization of systems and products broadly fall under two major categories such as MEMS and NEMS (Nano Electro Mechanical Systems). MEMS technology is considered as the amalgamation of two main sub-domains such as ultra-precision microengineering and IC (Integrated Circuit) technology. The scope of these domains is vast and considered to be disciplinary dependent. For instance, while the methodology of ultra-precision microengineering is more close to the mechanical machining and processes, IC technology in a generalized sense is correlated with microelectronics design. NEMS is again viewed as an interdisciplinary domain accommodating several disciplines, including general science! As is implied, MEMS and NEMS are miniaturized systems, which comprise electronics and mechanical components. In summary, it can be stated that MEMS and NEMS are both considered as interdisciplinary, multiphysics, multiengineering platform in which manufacturing of smaller parts, components, products and systems at the level of micro- and nanoscale with more functionalities and capabilities are easily realized.

The objective of this book is to provide detail information as far as proof-of-principle, concepts, design, development and applications of MEMS are concerned. NEMS will be presented in one chapter. It does this since MEMS has already been emerged as a fully-fledged technology, whereas development of NEMS is still at its rudimentary stage. NEMS groups of products and systems fall under a well-known technology called *nanotechnology*.

It is our intension to include most of the fundamentals covering the major topical subjects of MEMS and NEMS, however, it suffices to say at this point that one of the detrimental objectives of these technological advancements is being focused on manufacture and design of miniaturized systems that can meet the humane expectations to a satisfactory level. To some extent this chapter attempts to introduce a road map to the concept of MEMS and NEMS, which are explained in the following chapters in more detail.

In principle, both MEMS and NEMS should again be categorized under one engineering discipline such as Micromanufacturing, which is considered a sister discipline of Manufacturing. Both manufacturing and micromanufacturing disciplines are attributed under the branch of Production Engineering (PE). This book does not include the detailed topical subjects as far as PE is concerned, however, in order to grasp the subject matter at this stage of learning, a very brief description of important terms used in PE domain are outlined in the next section. Indeed, from the viewpoint of MEMS and NEMS the state-of-the-art PE and technologies is presented.

1.2 PRODUCTION ENGINEERING

Production engineering is the cornerstone of many industrial activities. It significantly contributes toward the economic growth of a nation. Generally, the higher volume of manufacturing of products in

a country, the better is the standard of living of the citizens. Production engineering is a process of making large quantities of products by effectively utilizing raw materials and related resources. It is a multidisciplinary design activity, which integrates both mechatronics engineering and manufacturing engineering.

1.2.1 Mechatronics Engineering

Mechatronics is defined as the synergistic integration of mechanical engineering with electronics and intelligent control algorithms in the design and manufacture of products. It deals with electronic controls of mechanical systems. Mechatronics is a relatively new discipline but it has firmly established itself. Technical areas such as motion control, robotics, automotive systems, intelligent control, actuators and sensors, modeling and design, system integration, vibrations and noise control are studied under this topical subject. It is an interdisciplinary engineering domain, which primarily considers control solutions of macroscale machine systems. Microscale machine systems incorporate MEMS methodology and can deal control solutions synergistically. The synergistic integration of MEMS technology and intelligent control algorithm in the manufacture of products is called *micromechatronics*.

1.2.2 Manufacturing Engineering

Manufacturing is the basis of production engineering. Manufacturing engineering is a design-process engineering that characterizes macro level design and especially production of mechanical products, although other items are included. It involves management information systems (MIS) and Computer Aided Process Planning (CAPP) that effectively enhance productivity. Manufacturing encompasses the design and production of goods and systems, using various production principles, methodology and techniques. They are Cellular Manufacturing (CM), Group Technology (GT), Flexible Manufacturing Systems (FMS), Just-in-Time (JIT) technology, Concurrent Engineering (CE) and Holonic Manufacturing Systems (HMS). The products vary greatly from application to application and are manufactured at several stages of their respective production processes. The product can be anything starting from a singular type such as automobile parts to complex type such as the automobile itself. The concept is, therefore, hierarchical in nature, in the sense that the process inherits cascade behaviour in which the manufactured product itself can be used to make other products or items. The manufacturing process may produce discrete or continuous products. Discrete products mean individual parts or pieces such as nails, gears, steel balls, beverage cans, and engine blocks. Continuous products are spool of wires, hoses, metal sheets, plastic sheets, tubes, pipes and so on. Continuous products may be cut into individual pieces and become discrete parts.

1.2.3 Scope of Production Engineering

The scope of production engineering is enormous. Figure 1.1 illustrates its scope citing MEMS and MEMS disciplines as the derivatives. Importantly, PE includes the following broad topics.

- Precision engineering and ultra-precision engineering
- Micromanufacturing
    - ◇ Microelectronics
    - ◇ Microelectromechanical Systems (MEMS)
- Nanotechnology

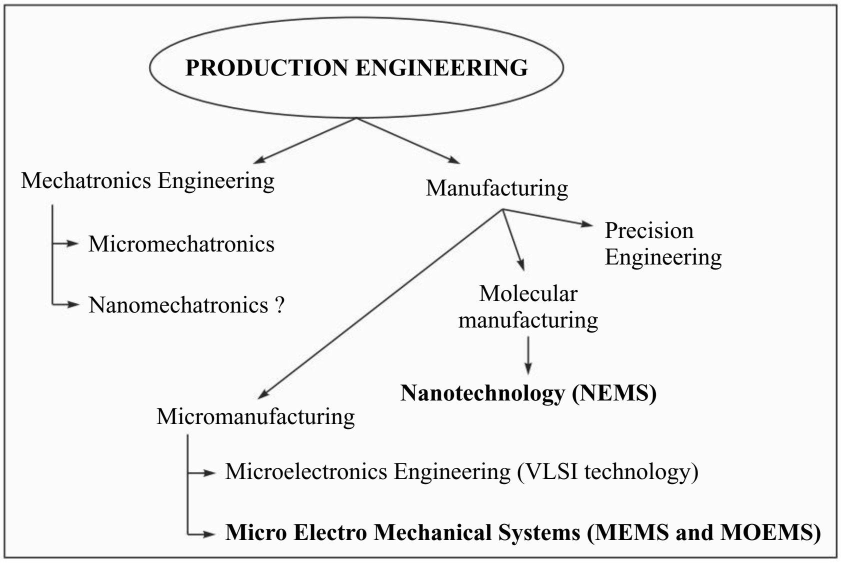


Fig. 1.1 *The branch of production engineering and its sister disciplines*

1.3 PRECISION ENGINEERING AND ULTRA-PRECISION ENGINEERING

The technical field of precision engineering has expanded over the past 25 years, although, in 1933, the Precision Engineering Society was established in Japan and thereafter the activities were accelerated with new impetus from Europe. The first issue of the journal Precision Engineering appeared in 1979 and the first academic program began in 1982 (Source: American Society of Precision Engineering (ASPE)). According to ASPE "…*the precision engineering* is dedicated to the continual pursuit of the next decimal place". The structural design, its modification and finishing methodology are the important topics studied under precision engineering. Further, precision engineering deals with quantitative assessment of the total uncertainty of the measurement. Surface finishing is a precision engineering activity. *Nanomachining* is an ultra precision engineering process. Nanomachining provides advanced material shaping in the nanometer realm. Note that many mechanical and optical parts in high-precision domain require precise structural perfection and manipulation at the nanometer scale.

Precision engineering includes design methodology, uncertainty analysis, calibration compensation, error compensation and controls. The main focus of this branch of engineering is metrology. A list with regard to application domains is given below.

- Instrument and machine design
- Dimensional metrology and surface metrology
- Materials and materials processing
- Interferometry
- Optics and scanning microscopes
- Semiconductor processing

1.3.1 Precision Engineering Process

Precision Engineering Process (PEP) is a concept of design, fabrication, and testing where variations in product parameters are caused by logical scientific occurrence. *Micromilling* and *microdrilling* are the

two important processes. Frequently used terms within PEP are scaling, accuracy, resolution and repeatability, which are known as performance measuring indices.

Scaling is a parameter that defines the ratio attributes with respect to the products to be prototyped or designed. It is considered as a fundamental attribute for predicting the behaviour of microstructures and systems for analysis and synthesis. *Accuracy* defines the quality of nearness to the true value. In the context of machine or production systems, accuracy is the ability of a machine to move to a desired position. As an example, if the actual value is 1.123 units and it is depicted as 1.1 units, we are precise to the first decimal place but inaccurate by 0.023. *Resolution* is a quality-measuring index. It is defined as the degree to which a change can be observed and detected. Several application-related examples could be given. In a servo control system, resolution is the fineness of position precision that is attainable by a motion system. In this example, resolution can be viewed in two ways: electrical and mechanical resolution. The smallest electrical increment that is produced by a servo controller is the electrical resolution. The smallest physical increment (minimum actual mechanical incremental move) that is achieved by the system is called mechanical resolution. Mechanical resolution is usually coarser due to the involvement of friction, *stiction* and deflections. *Repeatability* is the variation in measurements obtained when one person takes multiple measurements using the same instrument and techniques on the same parts or items. Repeatability is typically specified as the expected deviation, i.e. a repeatability of 1 part in 10,000 or 1:10,000.

1.3.2 Micromilling

Design of many three-dimensional (3D) miniaturized structures is based on micromilling process. Micromilling suggests two different approaches: cutting tool based approach and focused ion beam (FIB) based approach. Traditional cutting tool based design approach is characterized by milling tools which are usually in the order of hundreds of micrometers in diameter. Cutting tool based approach inherits shortcomings, as far as attainable desired values of scaling, resolution and accuracy is concerned. It should be pointed out that the forces present in micromilling with tools of the order of micrometer diameters are dominated by contact pressure and friction between the tool cutting edges and the workpiece. On the other hand, FIB overcomes the above drawbacks and, therefore, has become a very popular method. The specially designed milling machines, which utilize FIB based milling process, are called high-precision machines. The FIB process uses a sharp tungsten needle wetted with gallium metal. The tip of the needle is subjected to a 5–10 kV (sometimes higher) so as to enable the field ionization effect on the gallium. The gallium ions are then accelerated by the use of another energy source and focused into a spot of sub-micrometer order. The kinetic energy acquired by the ions makes it possible to eject the atoms from the work piece. This is a kind of *sputtering* process (described in Chapter 2). The sputtering yield varies inversely with the strength of the chemical bond in the materials. The movement of ions or the work piece, depending upon the situation, can be controlled to obtain wide varieties of 3D structures. As a rough calculation, one should note that in the FIB process, for a spot size of 0.45 μm with 2.5 nA of current, the required current density would be approximately 1.65 A/cm^2. Typically a high milling rate of 0.65 μm^3/nA s, corresponding to an average yield of 6.5 atoms/ion, can be obtained at 45 keV, 30° incidence, and 45 scans. The micromilling process is also applied for making micromolds and masks to aid in the development of microcomponents.

1.3.3 Microdrilling

Microdrilling is characterized by drilling of ultrafine holes. Drilling in the micro ranges, using the special microdrills requires a precision microdrilling instrument. The end of the microdrill is called the chisel

edge. The microdrills are made up of either micro-grain tungsten carbide or cobalt steel. Coarse microdrilling machines are available to drill holes from the size of 0.03–0.50 mm diameter in increments of 0.01 mm. However, presently the demand is of the order of micrometers. A sub-microdrilling technique utilizing the phenomenon of ultra-fast pulse laser interference (PLI) has been developed. For microdrilling and also for other delicate processing applications, Holo-Or Ltd. has developed an optical element that creates an output laser spot in the form of a top hat circle with a diameter of 350 μm. The PLI accepts a collimated Gaussian incident beam with a diameter of 12 mm from a 10.6-μm CO_2 laser. Even 300 nm holes can be successfully drilled on a 1000-Å-thick film using the interfered laser beam compared to micrometer holes using the conventional non-interfered laser beam. The most important parameters considered in the microdrilling are accuracy, sensitivity and quality. The applications of microdrilling are many. Some of which are as follows.

- Air bearings
- Bushings
- EDM tooling
- Electronic components
- Gas and liquid flow
- Microwave components
- Nozzles
- Optical components
- Semiconductor parts

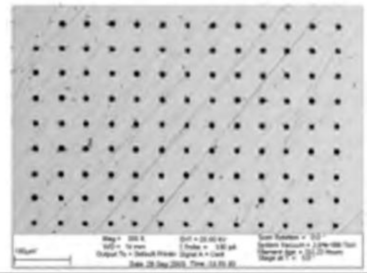
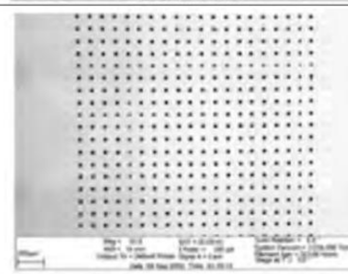
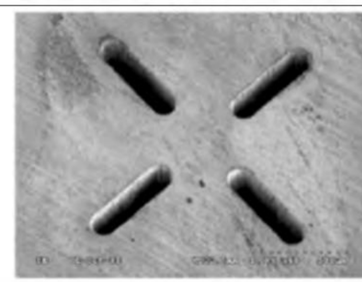
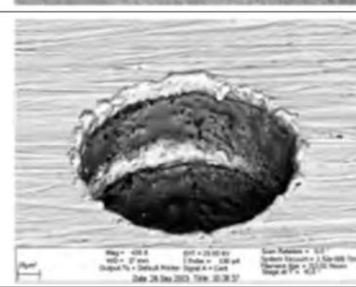
Aspect ratios up to 100 (e.g. a 15 μm hole in a 1.5-mm-thick foil) in thick material are achievable. High aspect ratio beyond 100 is practically not feasible. The major problem of laser microdrilling, however, is that they have short focal depth. The problem can be overcome by utilizing Bessel beam. Bessel beam can provide longer focal depth due to the fact that the beam is non-diffracting and practically they do not spread out. Bessel beam is generated using a pulsed laser in conjunction with relevant accessories. Example of microdrilling applicability using a typical laser system developed by ATLASER di Andrea Tappi, is illustrated in Table 1.1.

1.4 INTEGRATED CIRCUITS (IC)

For real-time operation, many electromechanical dynamic systems are interfaced with necessary electronic circuits, units and subsystems. The basic circuits, units and subsystems are arithmetic unit, logic circuits, registers and memory units, bridge circuit, Analog to digital converter (ADC) and Digital to analog converter (DAC), parallel-to-serial and serial-to-parallel conversion unit, transceiver, amplifier, filter and so on. All these circuits can be categorized into two groups; Analog circuits and Digital circuits. Analog circuits can be designed by using discrete components such as capacitors, inductors, transistors and diodes and digital circuits can be designed by using fundamental logic gates such as AND, OR and NOT. The design of necessary electronic circuits by using discrete components is a complex procedure and not reliable.

Instead of designing electronic circuits by using available discrete components, if we could fabricate the circuit within one cabinet at the manufacturing time, then probably, the complexity involved in connecting the individual components could be resolved. Indeed, it has been done through the design of what is known as Integrated Circuit (IC). ICs are Very Large Scale Integrated (VLSI) semiconductor chips, which are designed for intended applications and can replace a large circuit which otherwise would have been designed using discrete components. In fact Operational Amplifiers (OPAMP), Analog to Digital Converter (ADC), Digital to Analog Converter (DAC), Multiplexers, Encoders, Decoders, Latcher, Drives, etc. are some of the ICs available in the technology marketplace.

Table 1.1 *ATLASER di Andrea Tappi microdrilling system performance parameter*

Stainless Steel Sheet Thickness: 120 µm Hole diameter: 9 µm Hole pitch: 50 µm Matrix Process time: 0.15 s	
Si Wafer Thickness: 0.54 mm Hole diameter: 25 µm Hole pitch: 50 µm Process time: 0.65 s	
Silicon Carbide Wafer Thickness: 0.64 mm Through Hole: 130x500 µm In width: 130 µm Out width: 110 µm	
Cu-FR4 sandwich Thickness: 0.5 mm Hole dimension: 200 µm Process time: 3.3 s	
Aluminium Nitride Thickness: 425 µm In side width: 300 µm Outside width: 290 µm Drilling time: 33 s	

Courtesy: ATLASER di Andrea Tappi

1.5 MICROELECTROMECHANICAL SYSTEMS (MEMS)

Most of the phenomena are to be measured and subsequently used in a timely predictive manner in order to overcome realtime response limitations. The miniaturized systems have better response time, faster analysis and diagnosis, good statistical results, improved automation possibilities with decreased risk and costs. Therefore, considerable amount of research is being done to develop reduced version of existing systems down to micro and nanometer scale levels.

A technology that considers manufacturing of microscale, transducers, probes, capacitor, inductors, actuators, valves, gears, pumps, gyroscope, mirrors, switches, and so on, similar to semiconductor chips is referred to as microelectromechanical systems, or MEMS in short. In essence, MEMS are small and integrated devices, which combine electronics, electrical as well as mechanical elements (Fig. 1.2) to meet the control related functional requirements such as sensing and actuation. The size is in the order of micrometer range (Note that the size of a typical transistor is 1–50 μm). MEMS design technology is an extended form of traditional microelectronic IC fabrication techniques (Described later). Unlike microelectronic IC technology, MEMS technology can fabricate capacitors and inductors as well as mechanical elements such as springs, gears, beams, diaphragms, and so on. It was impossible to fabricate these components utilizing IC technology. IC technology can only fabricate conductors, insulators, and junctions (diodes and transistors). MEMS, therefore, is an advanced technology as far as micromanufacturing of microsystems are concerned.

Table 1.2 *Length scales and their corresponding units and symbols*

Length Scale	*Unit*	*Symbol*
10^{-15}	Femto	f
10^{-12}	Pico	p
10^{-9}	Nano	n
10^{-6}	Micro	μ
10^{-3}	Milli	m
10^{3}	Kilo	k
10^{6}	Mega	M
10^{9}	Giga	G
10^{12}	Tera	T
10^{15}	Peta	P

1.5.1 System-on-a-Chip (SOC)

MEMS is the integration of active and passive[1] elements on a single silicon substrate developed through more advanced IC processing technology. The active elements are sensors and actuators. The passive elements are passive electronic systems (PES) such as signal conditioning circuits (amplifier, ADC, filter, isolators, etc.) and passive mechanical systems (PMS) such as gear, crank, bearing, etc. PES, the microelectronic integrated circuits, can be thought of as the nervous system with sensors and actuators as the eyes and arms, respectively. While microelectronic circuits and components are designed by using IC fabrication process, the micro mechanical components are designed using *micromachining* process. MEMS promises to revolutionize most of the micro products by combing microfabrication

1 In electrical engineering and electronics, resistors, capacitors and inductors are usually called passive components. Active devices are based on semiconductor technology. These devices are based on the p-n junction. The diode (can rectify the input signal) and transistor (can amplify the input signal) are based on single and double junctions, respectively. Although a diode does not amplify, it is still considered an active device. However, the categorization of passive and active elements in MEMS domain is somehow different.

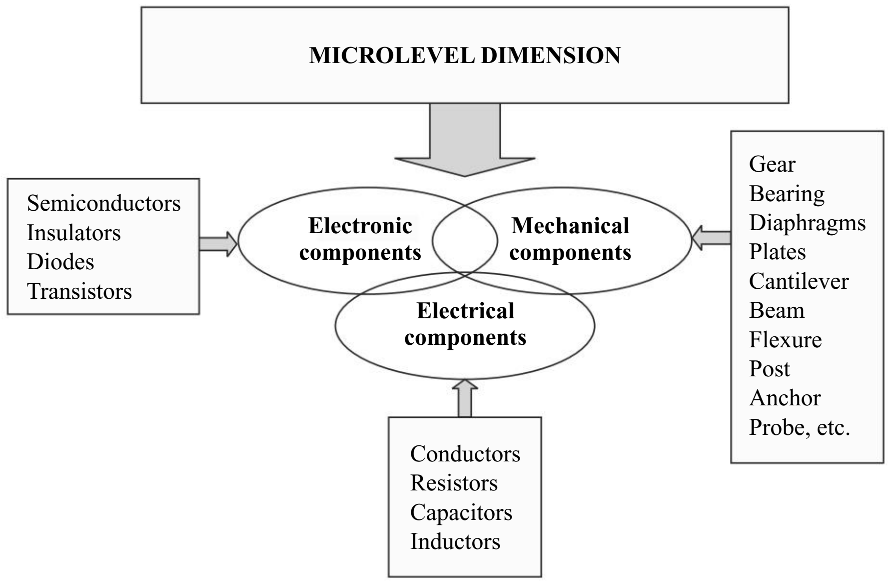


Fig. 1.2 *MEMS device comprises electronics, electrical and mechanical elements*

with micromachining processes sequence on silicon, making it possible to realize *systems-on-a-chip* (SOC) proof-of-methodology. Figure 1.3 shows a circuit model of an SOC device, which integrates electronics (amplifier and other passive elements) along with mechanical components such as resonators (described in Chapter 9).

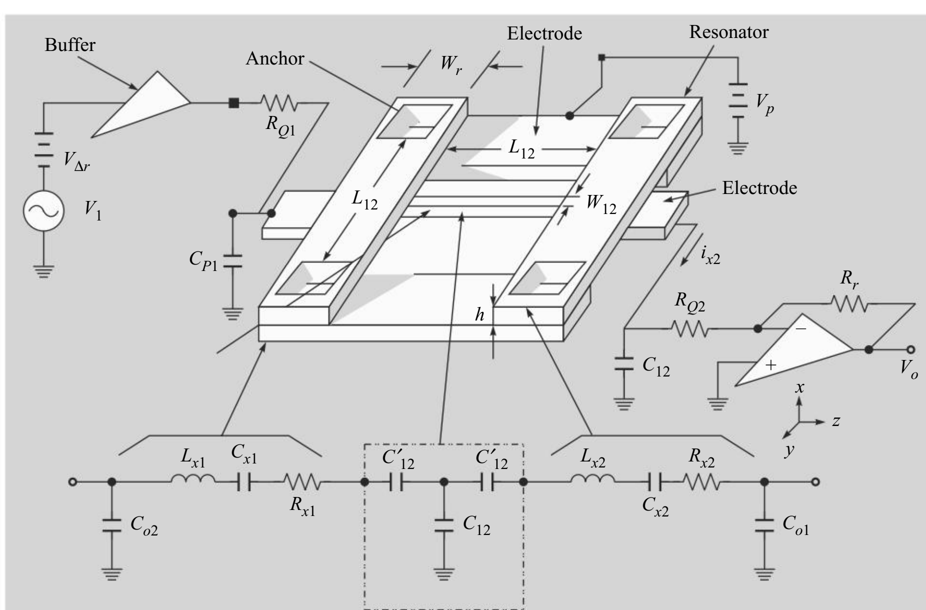


Fig. 1.3 *A typical system-on-a-chip RF resonator module*
Courtesy: Horizon House Publications, Inc., 2001

1.5.2 Scale of Integration

The density of MEMS chip is measured in terms of number of components required to make up a particular device. In other words the number of components present in the chip defines the *scale of integration*. Depending upon the scale of integration, broadly all the devices are categorized under four groups;

- SSI : Small Scale Integration
- MSI : Medium Scale Integration
- LSI : Large Scale Integration
- VLSI : Very Large Scale Integration
- VVLSI or ULSI : (Very VLSI or Ultra LSI)

SSI counts the components in the order of tens. MSI counts in the order of hundreds, LSI chips contains components in the order of thousands whereas VLSI integrates hundreds of thousand of components. VVSI incorporates more than what VLSI chip contains. Figure 1.4 shows the expected numbers of mechanical and electronic components in MEMS devices for typical applications as follows.

- MEMS for defense equipment
- Inertial navigation devices
- RF MEMS and optical communication systems
- Data storage applications

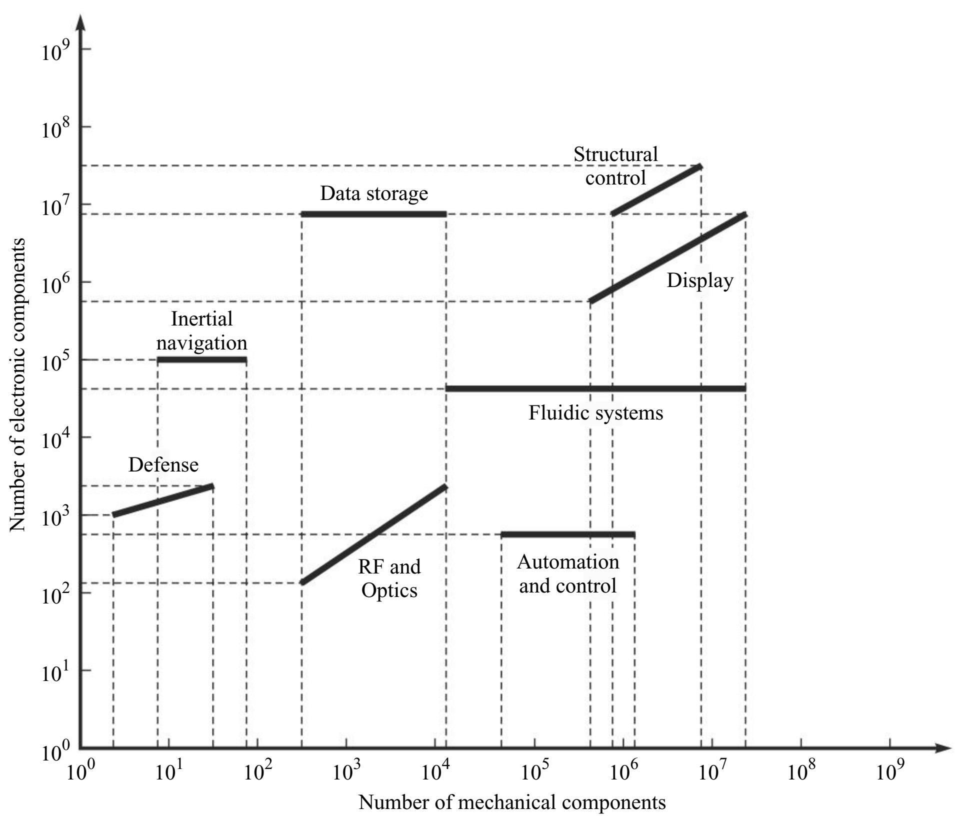


Fig. 1.4 *Density of typical SOC devices*

- Industrial automation and control sectors
- Microfluidic applications
- Display applications
- Distributed structural control

1.5.3 Next Generation MEMS

MEMS devices are manufactured by the use of batch fabrication techniques similar to those used for IC. Unparalleled levels of superiority, sophistication, functionality, reliability and availability can be achieved on a small silicon[2] chip at a relatively low cost. MEMS technology can allow the development of smart systems inheriting additional capability of perception and controlled attributes of microsensors and microactuators. The advanced system can lead to expansion of the scope of possible solution to reliable features for the target applications. For example, a quite large numbers of sensors can be micromachined in a single platform as a *sensor fusion device* (SFD). Sensor fusion is a scheme usually applied in the field of automation and control for reliable operation of control systems. Instead of using a single sensor, multiple sensors are employed in order to observe the same measurand. A single output however, is produced instead of many, based on majority voting. Since many sensors take part in generating a unified output, sensor fusion scheme can eliminate error(s) produced by the faulty sensor(s), within the group. The next generation MEMS device may also accommodate decision-making capability with embedded soft computing algorithm. Moreover, prognostic measure in terms of sensor and actuator validation can be achieved through MEMS integrated design approach.

1.5.4 Applications

MEMS devices have already found significant applications in many sectors. They are used for controlling micromanipulator, micro-handling equipments, microgrippers and microrobots. Many MEMS devices are found in clock, ink-jet printer head, color projection and display systems and scanning probe equipments. MEMS technology also designs many types of sensors including pressure, temperature, chemical and vibration sensors. MEMS-based light reflectors, beam splitter, RF and optical switches are common. Broadly the application sectors are:

- Aircraft industries
- Automotive
- Chemical, clinical and pharmaceuticals industry
- Automation industry and manufacturing sectors
- Defense and space applications
- Environmental
- Communications
- Health science (Pacemakers)
- Computing (Data storage devices, display, printing head)
- Consumer products

1.5.5 Market Growth of MEMS

Currently, MEMS market demands are becoming overwhelmingly high. MEMS-based systems developers are focusing on technological innovations, as they compete to offer products that meet customer requirements as well as performance. The market progressively is being strengthened by the fact that investment in MEMS is time-based value engineering that meets the high industry demand. User and vendors are also very sensitive to price of the market while formulating market strategies, at least over the short term. The value of MEMS products increased 14 billion USD by the year 2000, that was rightly predicted during 1995. The market for RF MEMS devices is forecasted to grow to 1 billion

2 Silicon has unique electrical, mechanical, optical, and chemical properties. Moreover, it is possible to integrate electrical, mechanical, optical, and chemical devices on a single piece of silicon crystal, i.e. a single silicon chip.

USD by 2006. Towards the beginning of 4th quarter in the year 1998, NEXUS (The Network of Excellence in Multifunctional Microsystems) task force announced the first 'Market Analysis for Microsystems', for the period 1996–2002. Their study included all types of microsystems, including MEMS. It estimated a projected market growth of 14 billion USD to 38 billion USD by the year 2002. Figure 1.5 shows the illustrations of market growth as far as sales of MEMS products are concerned. Figure 1.6 shows approximate percentage of various types of MEMS devices in various sectors.

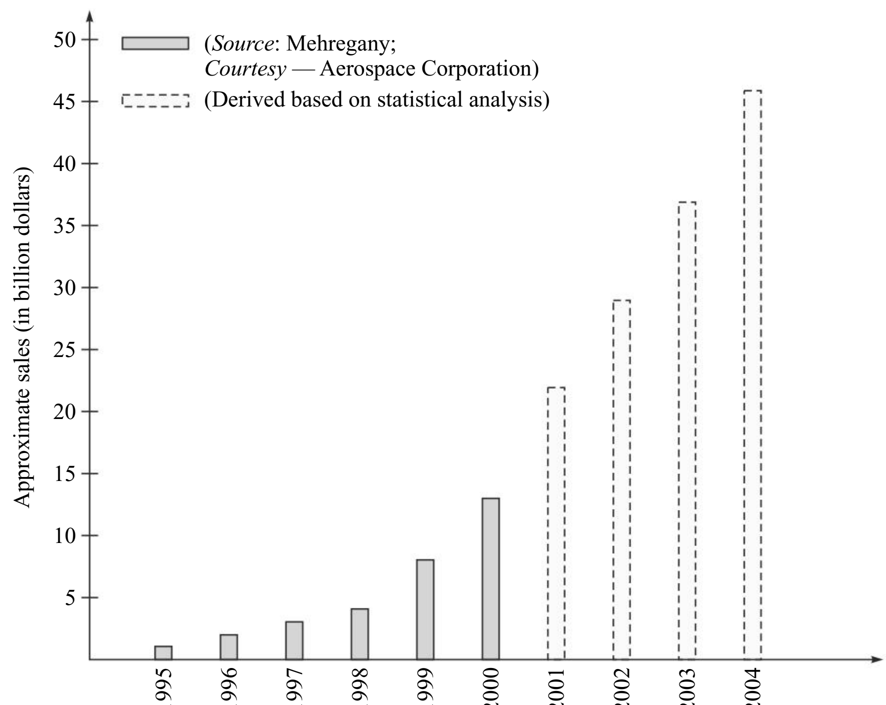


Fig. 1.5 *Market growth of MEMS based products*

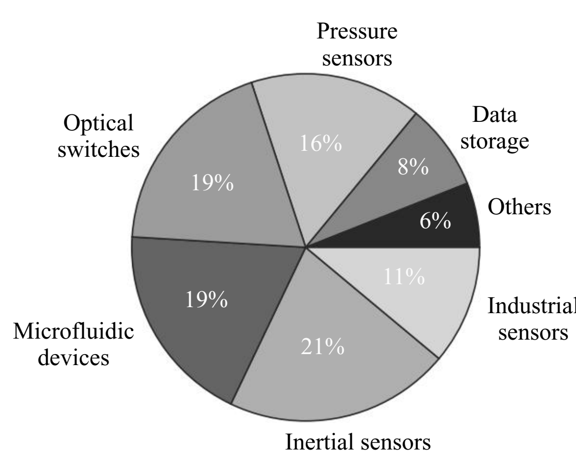


Fig. 1.6 *An approximate illustration of MEMS products in various sectors*

1.5.6 Topical Subjects

The core topical subjects under MEMS discipline are:

- Micromachining methods
- Sensing and actuating principles
    - ◇ Mechanical MEMS
    - ◇ Thermal MEMS
    - ◇ Magnetic MEMS
    - ◇ MOEMS and RF MEMS
    - ◇ Microfluidics and BioMEMS
- Simulation methods
- Performance measuring parameters

Before we introduce micromachining processes and follow up topics listed above some preliminary knowledge on sensors and actuators and their application domains is recommended.

1.6 MICROSENSORS

These sensors are used for measuring physical parameters. They are primarily employed to observe the temporal effects of the environment and subsequently to calibrate the observed values in order to produce meaningful information. The basic building block of a sensor is the transducer, which is a device that produces a measurable response to a change in a physical condition such as temperature, pressure, humidity, flow, light intensity, magnetic field, vibration, and so on. They respond to some properties of the environment differently, and transform the response into the electric signal, adopting various transduction principles. The important transduction principles are thermoelectric, photoelectric, electromagnetic, magnetoelectric, thermoelastic, pyroelectric, and thermomagnetic. MEMS sensors are designed with a broad range of transducers adopting these transduction principles.

The term transducer and sensor are used interchangeably, although there is a difference. Transducer is a basic element whereas sensor is a more sophisticated form of transducer with additional signal conditioning circuits. The signal conditioning circuits are interfaced at the output of the transducer. The circuits process the available raw signal of the transducer in order to produce a refined version. The signal conditioning circuit could be an amplifier, a noise eliminator, an offset compensator, an analog-to-digital converter (ADC) or combination of those circuits. In servo control systems sensors are used for providing feedback signals. In many cases the decision-making logics are also integrated into the sensor device. Wherever necessary the decision is then transmitted to the actuators, which respond to the command by moving, positioning, regulating, pumping and filtering actions. Thus, the transducer is an essential element of a sensor, which includes additional conditioning circuitry.

Sensors vary from very simple to highly complex. They are classified in different ways, such as according to the nature of signal they handle, selective material of the transducer, and the level of signal generation. One good way to look at a sensor is to consider the important attributes such as stimulus, specifications, physical phenomena, conversion mechanism, material, response, ruggedness, stiffness, range, ability to measure physical parameters, and application field.

Sensors, which are designed by utilizing MEMS technology, are called micromachined microsensors, or MEMS sensors. There are many applications of MEMS sensors. The list given below is of the MEMS based sensors used in a typical car. From this list one can imagine the types of sensors that are available in the technology marketplace to be used in other sectors.

- Air-conditioning compression sensor
- Force sensor
- Pressure and inertial sensor for braking control
- Inertial navigation sensor for sensing acceleration
- Tire pressure sensor
- Crash sensor
- Mass airflow sensor
- Exhaust gas sensor
- Microphone for noise cancellation
- Air-bag side impact sensor
- Fuel sensor
- Nozzle sensor for checking fuel injection

1.7 MICROACTUATORS

An actuator is a control device that makes something move. Many real-life technical systems require movement of some component of the target system at the microscale level. The process of micromovement is called microactuation. A microactuator accomplishes microactuation. In other words microactuator designed for typical target system produces the mechanical output (actuation). Microactuators are lightweight, conformable and precision devices. The microactuation applications are many, including:

- Actuation of micromirrors to scan laser beams
- Driving of cutting tools for microsurgical applications
- Driving of micropumps and valves for fluid and gas transportation
- Data reading and recording control
- Printing applications
- Spatial light modulation for display applications
- Optical signal switching
- RF signal tuning

One way to classify the microactuators is based on their movement. Accordingly, we have two types of actuations, (i) translational and (ii) rotational. A translational or angular movement of the order of micron, i.e. 10^{6} th of a meter or radians, respectively can be used for the applications listed above. It is possible to convert one form of actuation to another, for example, translational to rotational and vice versa, by suitably designing the interfacing components around the actuator. Just like a signal-conditioning circuit manipulates the transducer signal in order to produce a desirable output, the actuator output (movement) can be conditioned by suitably designing the interfacing components. The interfacing components include rigid body and passive mechanical systems (PMS). Anchors, gears, bearings, cams etc. can be considered interfacing components. Depending upon the load, range and environmental condition the microactuators are selected.

The basis of actuation is the method of energy transduction. Principal energy transduction methods, called driving methods, are mechanical, thermal, electrical and magnetic method. Accordingly, the commonly available microactuators are called:

- Mechanical actuators
- Thermal actuators
- Electrostatic actuators
- Magnetic actuators

Each driving principle has advantages and disadvantages with respect to power requirement, deflection range, required force, environmental durability, and most importantly the response time. Note that the power requirements depend mainly on the driving method involved. Table 1.3 illustrates the above four types of microactuators and their relative performance parameters in a generalized sense.

Of these, mechanical methods currently look promising, although the others have their places. Primarily mechanical driving method is based on piezoelectric driving principle. In a piezoelectric material, the internal dielectric displacement is developed by the applied electric field and mechanical stress. It depends on properties of crystals, ceramics and polymers. Thermal actuators require relatively large amount of electrical energy. The generated heat has to be dissipated. Electrostatic drive is based on

electrostatic coulomb forces that develop between capacitive-coupled microelectrodes that differ in voltage. When an external voltage is applied between the electrodes, a potential energy is stored that makes the actuation. The electrostatic forces act perpendicular to the parallel electrode. Electrostatic devices are also popular because they are relatively simple in structure, flexible in operation, and are fabricated from standard, well-understood materials. Electrostatic MEMS actuators that are investigated include parallel plate capacitor (PPC) based actuators, micromotors comb drive actuator and microvalve. Electrostatic actuation is possibly the most common and well-developed method, but it does suffer a little from *wear* and *sticking* problems. Electromagnetic actuation is primarily a current controlled process. The driving mechanism again requires currents of the order of several hundreds of milliamps and the voltages (driving) in the range of less than five volt. Magnetic drive is very suitable for the applications like dust-filled and conducting fluid environment. Magnetic actuators usually require relatively high currents (and high power). Magnetic and thermal actuators are sometimes called electromagnetic and electrothermal actuators, respectively, as both types require electrical current to produce magnetic field through conducting coil and heat through resistive materials, respectively.

Table 1.3 *Typical power requirements against driving principles*

Driving Method	*Power Usage*	*Voltage range*	*Current range*	*Force generated*	*Deflection range*	*Speed*
Mechanical	High	10–100 V	nA–μA	Moderate	Moderate	mS
Thermal	High	1–10 V	mA–10mA	Moderate	Small	mS
Electrostatic	Low	10–100 V	nA–μA	Moderate	Moderate	μS
Magnetic	Medium	1–5 V	~100 mA	High	Large	μS–mS

1.8 MICROELECTRONICS FABRICATION

One of the major invent in the last century is microelectronics. These are popularly known as microdevices. Microdevices can be IC chips that are designed at submicron dimension range. Microelectronics design methods entail accommodation of essential process attributes of micromanufacturing. The method of micromanufacturing of microelectronic devices is called *fabrication*. Fabrication is a sequential procedure. The important fabrication sequences are wafer preparation, film growth, doping, lithography, etching, dicing and packaging. All these processes are performed on a wafer called substrate. A *substrate* is a bulk of semiconducting material. A polished silicon crystal is commonly used as the substrate. A long single substrate can accommodate to fabricate several identical chips. Hence, IC fabrication process is a batch-processing scheme.

On the substrate a thin film (layer) is grown. Then the properties of the layer are modulated by appropriately introducing *doped* material in a controllable manner. Two commonly used doping methods are *thermal diffusion* and *ion implantation*. The subsequent process is called lithography, which refers to creation and subsequent transformation of a masking pattern. A *mask,* which is prepared beforehand, consists of a glass plate which is coated with a patterned layer, usually chromium film. The preparation of mask can be carried out by the use of computer-controlled electron beam to expose the photographic mask material in accordance with the desired configuration and requirement. The configuration information is supplied to the computer. Mask represents the features of the various elements and layers of the chip to be manufactured. The pattern on a mask is then transferred to the film

by means of a photoresist. *Photoresist* is a chemical and light sensitive that is coated on the top of the thin film already grown. The mask features are transferred to the film of the substrate by exposing the light-sensitive photoresist coating through the transparent areas of the mask. The mask then appears on the surface of a thin layer. Once the pattern is transferred into the film, the material areas of the substrate unprotected by the hardened photoresist are then removed by etching. The transformation of pattern is called lithography. When UV light is used it is called photolithography. X-ray and electron beams are also preferred.

Etching is a process of removing the portions of a layer by chemical or electrolytic means. Etching can be either physical or chemical, or a combination. There are broadly two types of etching process: wet and dry etching. In wet etching, the material is dissolved by immersing in a chemical solution (known as etchants), on the other hand in dry etching, the material is dissolved using reactive ions or a vapor phase *etchant*. Some of the commonly used etchants and their etch rate with respect to Si(100), are given in Table 1.4. Etching techniques are characterized by their selectivity and degree of anisotropy (more details in Chapter 2). After the etching process, a thin aluminium or gold layer is deposited on the uppermost layers of the chip in order to allow metal to contact the device elements. The metal layer interconnects the elements and also provides a means to external electrical connections. The aluminium deposition is often achieved by using chemical vapor deposition (CVD) method. *Dicing* is a process of cutting up the wafer into individual chips. *Packaging* is a complex process that involves physically locating, connecting, and protecting a device.

Table 1.4 *Characteristics of major wet etchants*

		Etch rate			
Etchant	*Suitable masks*	*Si(100) (µm min^{-1}) (100)/(111)*	*SiO_2 (Å h^{-1})*	*SiN (Å h^{-1})*	*Comments*
Hydrazine	SiO_2, SiN Most metals	0.5–3 16:1	100	<<100	Highly toxic Explosive at high conc.
EDP	Au, Cr, Ag, Ta, SiO_2, SiN	0.3–1.5 35:1	120	60	Highly toxic
KOH	SiN,Au	0.5–2 up to 200:1	1700–3600	<10	Not cleanroom compatible
TMAH + IPA	SiO_2, SiN	0.2–1 up to 35:1	<100	<10	Cleanroom compatible

Source: French et al., *J. Micromechanics and Microengineering*, 8, 1998, 45–53

The design of an IC with millions of transistors and even more interconnections is not a trivial task. The fabrication is considered a technology. As mentioned earlier it utilizes computers during the development process. Prior to its real design the circuit is prepared and tested using EDA (electronic design automation) Tools. EDA tools are computer assisted development (CAD) tools. These tools help in synthesizing and simulating the behavior of the desired circuit in terms of arranging the placement of transistors and interconnections within the area of the chip. It can also verify and validate all defects and conditions, respectively. The microelectronics fabrication technology has been driven by the demands of computer industry, space technology, car industry and telecommunications.

1.9 MICROMACHINING

The term micromachining refers to the fabrication of 3D MEMS structures with the aid of advanced lithography followed by etching. Lithography patterns the structural material whereas etching removes the selective portion of the substrate or thin film based structural and sacrificial material already deposited. In general, the micromachining process can either use the material to form microstructures by etching directly into the material or use structural sacrificial layer to produce the same. Sacrificial layer is etched away in order to obtain a freestanding 3D structure. Broadly, the fabrication processes fall into two categories such as:

- Bulk micromachining
- Surface micromachining

1.9.1 Bulk Micromachining

Bulk micromachining refers to etching through both the sides (front and back) of a bulk of material to form the desired structures. The structures are formed by wet chemical etching or by reactive ion etching (RIE). Usually, suspended microstructures are fabricated using wet chemical etching. The advantage of bulk micromachining is that substrate materials such as quartz or single crystal silicon are readily available and reasonably high aspect-ratio structures can be fabricated. It is also compatible to IC technologies, so PES (passive electronic systems—see Chapter 4) can be integrated, easily. The disadvantages of bulk micromachining is that the process is pattern and structure sensitive and pattern distortion occurs due to different selective etch rates on different crystallographic planes. Further, since both frontside and backside are used for processing, severe limits and constraints are encountered on the minimum feature size and minimum feature spacing.

Bulk micromachining utilizes the etch selectivity between {111} planes and {100} and/or {110} planes in aqueous alkaline etchants. Si wafers with (100) and (110) orientations are essentially used in bulk micromachining. Using (100) silicon, simple structures such as diaphragms, V-grooves, nozzles, and more complex structures such as corner cubes and rectangular masses, can be fabricated. Vertical microstructures can be fabricated on Si(110) wafer. It is because Si(110) has four {111} planes that intersect the wafer surface vertically.

1.9.2 Surface Micromachining

Surface micromachining is another method that characterizes fabrication of MEMS structures out of deposited thin films, i.e. it involves the creation of mechanical structures in thin films already grown on the surface of the wafer. Layers from which the 3D structures will be created may be composed of three layers as follows.

- *Isolation layer:* When silicon substrate is used as the substrate, the first step in surface micromachining is the deposition of an isolation layer. This layer is deposited with dielectric material such as silicon oxide (SiO_2) followed by a thin layer of silicon nitrite. This acts as etch stop for many etchants.
- *Sacrificial layer:* Sacrificial layer is also called spacer layer, which needs to be etched in order to create freestanding 3D structure. A phosphosilicate glass (PSG) layer is a preferred material for sacrificial layer.
- *Structural material:* This is the layer from which the structure will be built. The most common structural material in microfabrication is polysilicon (poly-Si or simply poly). Polysilicon material

based surface micromachining has been the backbone of the fabrication technology for many of the microsensors and actuators.

Silicon has excellent mechanical properties making it an ideal material for machining. The layers are deposited in sequence and subsequently some selective portions of the sacrificial and structural layers are removed to build up a 3D mechanical structure. Hydrofluoric acid can dissolve the sacrificial layer. Rinsing and drying follow etching. Once done, the structure can be freed from the planar substrate. This is called release process.

The surface micromachining process is a critical method, as it requires serious attention as the property of material significantly varies at the microstructure level. In particular, following issues are dealt with careful attention.

- Basic understanding and control of the material properties of structural films
- Releasing method for the microstructure
- Fabrication features for hinged structures and high-aspect ratio devices
- Packaging methods

1.10 MECHANICAL MEMS

Mechanical MEMS mostly emphasizes two classes of devices; mechanical structure based device and piezoelectric material based device. When the geometric structural configurations are exploited for sensing and actuating purpose then the MEMS design can be classified under the first category. Various geometric structural configurations are cantilever, beam, plate, diaphragm and hollow chamber. Piezoelectric material based mechanical sensors and actuators exploit the effect of piezoelectricity. As piezoelectricity concerns with mechanical properties such as stress and strain, piezoelectric material based MEMS devices have been classified under mechanical MEMS. Broadly, the mechanical MEMS utilize the following methods and principles.

- Cantilever beam as sensing element
- Use of plates/diaphragm structure for capacitive sensing
- Microphones as sound sensor
- Exploitation of Coriolis acceleration in gyroscopes for angular rate measurement
- Principle of piezoelectricity and piezomechanics

Cantilevers bend when pressure is applied and oscillate in a way similar to a spring. Cantilever sensors can be used for the detection of physical, chemical and biological analytes with relatively good sensitivity and selectivity. Capacitive sensing has many applications, which will be explained in the following chapters. The application areas are vast including acoustic measurements, vibration monitoring, viscosity and density measurement, infrared and UV radiation detection, magnetic and electric field detection, detection of chemical vapors, including medical and biological agents, measurement of contaminants in water, explosive vapors, nuclear radiation and detection of DNA.

A microphone is an electro-mechanical-acoustic transducer that transforms acoustical energy into electrical energy. These are air-coupled ultrasonic microsensors, which take the advantages of miniaturization and low power consumption offering a wide range of applications such as sound detection and analysis, wind noise flow turbulence sensing and vibration sensing. The principle of operation of such sensor is based on the mechanical vibration of micro-membranes or diaphragm. The diaphragm is a thin, circular membrane that makes up a capacitor with the ground plane. The value of

the capacitance changes during the vibration cased by the sound signal. The deformation or deviation of the membranes from the normal value depends on the amplitude of the incident pressure (sound pressure).

Measurement of the angular rate of rotation is useful in many applications. A very common application is the measurement of the orientation or tilt of a vehicle running at high speed in a curved path. MEMS gyroscopes are designed to measure angular rate of rotation. The gyroscope exploits the Coriolis acceleration through a microplate with rotation-induced lateral deflection being sensed capacitively.

Piezoelectric materials such as lead zirconate titanate (PZT) are promising materials for MEMS applications due to their unique and remarkable properties. The PZT convert mechanical disturbances to electrical signals. Piezoelectric polymers are also now being used profusely as they offer the advantage of strains without fatigue. Many MEMS accelerometers employ piezoelectric sensing technique, utilizing cantilever beams. Reverse of piezoelectricity is called piezomechanics which can be exploited for the design of piezoactuators.

MEMS mechanical sensors are very popular because of easy integration procedure in the micromachining process. The basic challenge encountered in designing, however, is the implementation of signal processing circuitry. Further, it inherits many disadvantages which are given below

- The overall silicon area is generally larger
- Larger volume with respect to packages
- Multi chip modules require additional integration steps
- Existence of larger signals from the sensor output
- Stray capacitance of the interconnections

1.11 THERMAL MEMS

Thermal MEMS work on thermal phenomena. Thermal phenomena are basically described in terms of thermoelectric effect, Peltier effect, thermoresistivity, pyroelectricity and shape memory effect.

German physicist Seebeck discovered the thermoelectric effect. The phenomenon shows that a voltage (current) is developed in a loop containing two dissimilar metals, provided the two junctions are maintained at different temperatures. The effect is popularly known as Seebeck effect. The reverse Seebeck effect is called the Peltier effect, which states that in the presence of a current a temperature difference between the two junctions is produced. Thermoresistivity refers to the variation of electrical resistance (or conductance) of metals and semiconductors with respect to temperature. The variation of resistance in the material, called thermo resistive material is exponential. In principle, the conductivity of some material is proportional to its free charge carrier (FCC) density. More FCCs are liberated by thermal energy. Thermoresistive materials, also called thermistors, are of two types: negative temperature coefficient (NTC) and positive temperature coefficient (PTC) thermistors. In the former case, resistance decreases with increase in temperature and the reverse is true in latter case. Pyroelectricity is the migration of positive and negative charge to opposite ends of the polar axis of some typical crystals as a result of a change in temperature. Some of the important pyroelectric materials are lithium tantalate ($LiTaO_3$), Gallium nitride (GaN), triglycine sulfate (TGS) and polyvinyl fluoride (PVF_2). Pyroelectric materials are very sensitive. For instance thermal radiation of human being is sufficient to create measurable electric voltage. This is widely used for motion detectors. The shape memory effect is the ability of some specific alloy to recover the original shape under heating. Initially, the alloy is pre-

strained in low-temperature enabling it to deform severely from its original shape. By applying heat flux the original shape of the alloy can be recovered. The alloy is popularly known as Shaper Memory Alloy, SMA in short. When SMA exhibits the shape memory effects upon both heating and cooling, then it is called Two-way Effect. SMA has two solid-state phases; Martensite and Austenite phase. Martensite phase is relatively soft and exists at lower temperatures. At this phase, the alloy can be easily deformed. In martensite phase the alloy can remain at undeformed state or at deformed state. Loading or pre-straining can deform the alloy. On the other hand, austenite phase is the stronger phase of SMA and occurs at higher temperatures. The temperature at which an SMA undergoes a phase change from austenite to martensite is called Memory Transformation Temperature (MTT). The most effective and widely used alloys are NiTi (Nickel-Titanium), CuZnAl and CuAlNi. The topical subject under thermal MEMS includes:

- Thermodevices including principle of thermocouple and thermopiles
- Peltier heat pump and heat sink devices
- Hotwire and microhotplate based thermal flow sensor
- Application of micro-thermo-vessels
- U-shaped horizontal and vertical uni- and bidirectional thermal actuators
- Bistable MEMS relays for power system applications
- Chevron actuator
- Thermocouple probe for imaging, topography and data storage applications

1.12 MOEMS

Micro Opto Electromechanical Systems is abbreviated as MOEMS. MOEMS are MEMS, but they handle optical signals instead of traditional electrical signals. MOEMS technology accommodates the principles of optics, electronics and mechanics. MOEMS technology requires a different set of rules for operation as opposed to normal MEMS world. MOEMS show good performance with negligible signal degradation and better quality of service (QoS) compared to traditional optoelectronic devices. High operational bandwidth and low power consumption are the key features of MOEMS devices. The advantages of MOEMS are as follows:

- The speed of operation is high
- High bandwidth: Can handle many signals simultaneously
- Insensitivity to electromagnetic interference
- Can be used in harsh environments
- Secured and reliability

MOEMS have emerged to provide unparalleled functionality in telecommunication applications. Manufacturers of these devices are forecasting new opportunities in information technology, health care, military, industrial, and test and measurement sectors. Some of the important applications of MOEMS are listed as follows.

- Free-space optical switches, routers and beam splitters
- Focusing components
- Tunable filters
- Display and projection systems
- Guided optic devices and tuners

One very important combined process that is being used in micromanufacturing the MOEMS devices is the LIGA process. LIGA is a German acronym for lithography, electroplating, and molding. The process states that each layer of the different material is deposited lithographically and these layers are different thicknesses and can overlap one another depending upon what the MOEMS device is being designed for. One layer is usually used as a sacrificial layer to fill in and support a void area. The subsequent layer can be used to make a mold for the next layer. The succeeding layer will overlap the first and be molded into shape by the second. Then the first two layers can be removed. This in turn leaves a freestanding structure.

The topical subjects of interest are

- Properties of light and their exploitation with respect to MOEMS
- Optical switching principle, concept, design, and applications
- Beam splitting and microlens fabrication
- Principle of Spatial Light Modulation
- Light detector or wavelength separator using micro-optic waveguide (MOW)

1.12.1 Spatial Light Modulator

A Spatial Light Modulator (SLM) is a device that modulates light according to a fixed spatial pixel pattern. SLM plays vital role in several optical areas where light is controlled on a pixel-by-pixel basis. The general idea of such devices is to modulate optical wave front as a function of input signal. It consists of an array of electrostatic parallel plate actuators, which are directly coupled with square-shaped silicon based tiltable mirror pixels. The tiltable mirrors have the potential to modulate spatial and temporal features of an incoming optical wavefront, and have wider applications in imaging, beam-forming, and optical communication systems. The tilting of the mirror is achieved either by mechanical deformation or Coulomb force of attaction. The latter method is preferred and it takes places as a result of the electric field imposed by potential differences between the addressable electrodes. The SLM are very fast, compact and reliable devices with excellent optical quality as compared to the traditional electron gun based display technology. The devices are fabricated using a three-layer polysilicon surface micromachining process. Details will be presented in Chapter 7.

1.13 MAGNETIC MEMS

Magnetic materials play important role in designing MEMS sensors, actuators and storage devices. The magnetic materials could be soft or hard. The use of magnetic materials in MEMS is a recent development. Soft ferromagnetic materials have found the most utility in microsensors, microactuators and microsystems. Hard magnetic materials have several applications including storage devices. Permanent magnets are used for sensor and actuator applications that can provide a desirable constant magnetic field without the consumption of electrical energy. The other important characteristic is that the permanent magnets do not generate heat. The energy stored in a permanent magnet may not deteriorate if the magnetic material is properly micromachined. Long-range force and deflection with low driving voltage can be generated.

Magnetoresistive (MR) materials are used for sensing applications. They are used for detecting the strength and direction of the magnetic field, which in turn can measure the distance, proximity, position, angle and rotational speed. The MR materials undergo a change in resistance in response to an applied magnetic field vector. The variation in resistance depends on the rotation of magnetization relative to the

direction of the applied field. The change in resistance of the material is highest if the magnetization is parallel to the field and lowest if it is perpendicular.

Permalloy has good soft magnetic properties, high permeability, high magnetoresistive effect, low magnetostriction, stable high frequency operation, and excellent mechanical properties. In hard-disk magnetic recording heads, permalloy is widely used. Devices such as magnetic separators, micropumps, magnetic micromotors, inductors, switches, and microrelays have also been fabricated using permalloy as the magnetic material.

Actuators are capable of generating large bi-directional forces with long working lengths. A number of micromachined magnetic actuators require materials with desirable magnetic properties such as high permeability and thicknesses in the range of several micrometers. The actuator can be uni- or bidirectional type depending upon the way the actuation takes place. As name suggests the bidirectional actuators can drive in positive and negative direction with respect to reference. Deflections of over 1000 micron can be achieved with surface micro machined cantilevers, magnets and pancake coils (described in Chapter 8). A typical magnetically actuated microcantilever incorporating a magnet is measured 2000 by 1000 by 100 microns. Deflection at the tip of the cantilever can be about 1000 microns subjected to a driving current of 100 mA passed through a 40 turn pancake coil. The actuation force depends on magnetic field intensity, the magnetic susceptibility, and the mass of magnetic susceptible material.

The following topics are covered with regard to principles, concepts and application of magnetic MEMS technology.

- Magnetic materials, their properties and applications
- The principle of magnetization and magnetoresistivity
- Magnetic sensors and actuator fundamentals
- Magnetodiode and magnetotransistor
- Proof-of-principle magnetic pressure sensor and RF switches
- Bi-directional actuator fundamentals, operation and fabrication
- Typical application of magnetic actuator for microsurgery applications
- Large force integrated VR actuator
- Magnetic probe based storage device

1.14 RF MEMS

Wireless communications have existed for a long time. New wireless communication systems are being developed more rapidly than ever. Wireless technology utilizes RF (Radio Frequency) signal, which is an electromagnetic (EM) signal. Radio frequency operates in the range 9 kHz to 300 GHz; however Table 1.5 shows some frequency range of interest. RF MEMS is an emerging technology that plays a major role in accelerating the current growth in wireless communication. The impact will be felt at all levels in the wireless infrastructure from high-end transmission stations to low-end consumer

Table 1.5 *Some frequency ranges of interest*

Frequency	*Application*
2 kHz	The human voice
530 kHz	AM Radio
54 MHz	TV Channel 2
88 MHz	FM Radio
746 MHz	TV Channel 60
824 MHz	Cell Phones
1.85 GHz	PCS Phones
2.4 GHz	Wireless LANS
4.2 GHz	Big Dish Satellite
9.0 GHz	Radar
11.7 GHz	Small Dish Satellite

products, especially mobile phones. RF MEMS add new capabilities and improved power efficiency, while keeping wireless devices small and affordable. Broadly, the components and devices that are used for RF communication are switches, inductors, varactors, filters, tuners and resonators. The important components of a typical mobile phone are inductors, variable capacitors, RF switch and resonator. MEMS versions of these components promise to make systems more reliable and power efficient. From Fig. 1.7 one can observe how a mobile handset comprises RF MEMS based devices and modules. RF MEMS can be used for achieving:

- Transmission and reception
- VCO tuning
- RF band select filters
- Intermediate Frequency (IF) filtering
- Time delay for phased-arrays
- Variable Delay Lines (VDL)
- Reconfigurable antennas design

Fig. 1.7 *MEMS components in a typical mobile phone*
Courtesy: Sony Inc.

1.15 MICROFLUIDIC SYSTEMS

The study of transportation of fluids and their mixtures at a microscale level is known as microfluidics. Microdevices, which are used to transport and store fluid are called microfluidic systems (MFS). Typically the MFS handle fluid volumes in the order of nanoliter. Some of the important building blocks of microfluidic systems are:

- Microchannel
- Micronozzles
- Micropumps
- Microvalves
- Microreservoirs

There are a great number of applications of microfluidic systems. Some important applications are inkjet printing, drug dispensing, reaction analysis, mixing and separation, chemical synthesis, detection of chemical species, genetic analysis and semiconductor processing. The advantages of MEMS compared to conventional fluidic systems are that the miniaturized system requires less reagent (species or samples) resulting in faster, accurate and reliable measurements. Overall, the main advantage of MFS is better performance.

MFC requires construction and design that significantly differs from macroscale hardware as the behaviour of fluid at microlevel is somehow different. For instance, the capillary action changes significantly when the fluids pass through microscale diameter channels. As the scale becomes smaller, the dimensions of a device reach a certain size and the fluid particles or the solvent become comparable in size with the channel or the device itself.

MFC are widely used in semiconductor processing technology. The requirement of MFS in semiconductor industry is for gas distribution and control. MEMS-based technology can create pressure regulators, shut-off valves, and mass flow controllers (MFC) for electronics specialty gases (ESG) distribution. The use of MFS and modules eliminates the size and number of welds and face seals used in integrated gas control and distribution components. Regardless of the application domains of MFS, the technological design issues to be addressed are as follows:

- Precision alignment, accuracy, geometrical regularity and smoothness
- Mechanical parameters such as resistance to moderate and high pressures
- Architecture for complex structure and packaging density
- Standardization issues

1.16 BIO AND CHEMO-DEVICES

Microdevices used for analysis and detection of biomedical and industrial reagents are called bio and chemo-devices. Unlike MFS, bio and chemo-devices are diode-type, capacitor-type, transistor-type or 3D cantilever structure. The devices for sample analysis for biomedical and industrial needs are still under development. Some of the applications are

- Forensics
- Genetic screening
- Stress-response analysis
- Antibodies gene expression in transgenic cells
- Bio-warfare agents detection
- Bacteria detection
- Drug discovery, analysis and synthesis

The development of DNA sensors is considered as the most innovative in molecular biology technology. These micro or nano systems allow an easy, accurate, fast and reliable way to analyze many DNAs simultaneously. One innovative development is the eSensor™ DNA Detection System by Motorola Inc. using which reliable DNA testing can be achieved. The living being uses DNA to store its genetic code that describes everything about the organism. Specific fractions of the human DNA reveal valuable data. This data can be utilized for health care and medication purpose.

The basic element of the DNA analysis system is the DNA probe(s). The technology is based on the recognition of oligonucleotides[3] by their complementary genomic sequences[4]. The process of recognition is called hybridization. Hybridization can be achieved in many ways such as optical method,

3 An oligonucleotide corresponds to a very short DNA single strand

4 The relative order of base pairs, whether in a DNA fragment, gene, chromosome, or an entire genome

surface stress and electrical detection. Optical method is a traditional method, in which the color of the solution, in some cases change. Surface stress method uses a probe of micro cantilever probe. The probe is coated with the detector film that reacts with the biomolecules of interest. A biochemical reaction at the probe surface changes the surface stress. The surface stress causes a bending of the cantilever. A piezoresistor integrated with the cantilever sense the bending. The bending curvature is the measure of the presence of DNA. The electrical characteristic based process uses microelectrodes. The electrical resistance between the electrodes is infinite. The DNA probes are attached to electrodes. A solution of unknown sequence (target) DNA strands is then brought. Compatible DNA strands bind to the attached probes and the unreacted strands do not. Any DNA that hybridizes with the probe is the DNA to be detected. Upon hybridization, a bridge like structure (Fig. 1.8) is formed on the electrodes. Because of the formation of the bridge the electrical resistance between the electrodes drop from infinite to few thousand ohms. The change in resistance value is the measure of presence of DNA. The electrical characteristic based detection of a specific DNA molecule is very easy compared to surface stress based method.

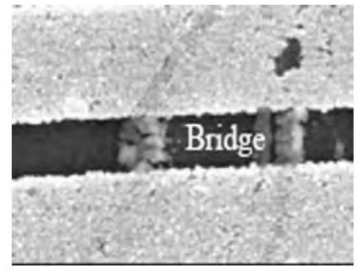


Fig. 1.8 *DNA hybridization*
Source: Fuller, Rochester Institute of Technology

1.17 NANOTECHNOLOGY

Nanoscale devices and equipments can give added benefits in terms of further miniaturization, efficiency and resource consciousness. Nanotechnology has accelerated research and development. It finds abundance applications in automotive, aerospace, household appliances, sporting goods, telecommunications equipment, and medical.

Carbon is suitable for the nanotechnology-based products. Pure carbon exists in four different crystalline forms: Diamond, Graphite, Fullerenes and Nanotubes. Nanotubes can be utilized as electronics devices, super-capacitors, lithium ion batteries, field emission displays, fuel cells, actuators, chemical and biological sensors. Much of current researches focus on potential applications of carbon nanotubes (CNT). Carbon nanotubes are considered as the ultra-fine unique devices having remarkable mechanical, electronic and chemical properties. CNT typically are longer in length. The length and diameter of the tube ranges from about a few tens of nanometers to several micrometers and 3–30 nanometers, respectively (Diameter is about 10,000 times smaller than a human hair). Each nanotube is a single molecule made up of a hexagonal network of covalently bonded carbon atoms. Freestanding carbon nanotubes can be grown by chemical vapor deposition (CVD). They possess metallic to semi-conducting properties along with good thermal conductivity. Other desirable properties they possess are high tensile strength, resilience and current densities.

Various nano level structural constructs can be formed through appropriate methods. X-like junction with diverse angles between the branches can be formed. One of the branches of the X-junction can be removed, creating Y- and T-like junction. These tubular junctions exhibit electrical transport behavior. Different junctions show different current versus voltage characteristics giving rise to the possibility of using as nanoscale electronic devices for *nanocomputers*. Nanotube-based design scenario even expects the design of gears and bearings and hence the machines at the molecular level.

A fuel cell is an electrochemical system that generates electricity by converting chemical energy into electrical energy. The fuel cell can be thought of as high-efficient next generation micro power system. Several challenges to widespread implementation of fuel cell technology are needed, although novel inexpensive and long-lasting electrocatalyst materials are becoming major factors in design and development. Carbon-based materials have suggested the possibility of using carbon nanotubes as novel electrocatalyst supports.

1.18 MODELING AND SIMULATION

MEMS devices are designed using micromachining technique. Prior to their design, it is desirable to study the behavior of the systems.

Consider a parallel-plate capacitor type electrostatic MEMS actuator. If a voltage were applied across the two electrodes of the microactuator then the movable plate would be displaced resulting *actuation*. When an external forcing function is applied to the system a change may not occur at all. The change occurs only when the input overcomes the loss component that is pragmatic in a real system. The way the system responds to input depends on both input forcing function and on the other factors such as its geometry and the material using which the system has been built. The law of conservation of energy always governs the relationship between the input and the output of a system. In this example the external forcing function was voltage and the system was the capacitor. In another example, the external function could be a force, pressure or magnetic field. What is important is that each and every system can be seen through some form of governing equations, which are formulated based on input and output relationships. The governing equations are called model equations. In other words, the model equation represents the system. The designer analyses system behaviour (observation of output in response to input) from the model equations. If model equations are given it can be used straight away, else the equations have to be formulated simply by looking at the system. This is called identification. Capacitor-based actuator is considered a simple dynamic system and its analytic dynamic model equations can immediately be written as,

$$C = \varepsilon A/l(t) \tag{1.1}$$

$$F = Q^2(t)/2\varepsilon A \tag{1.2}$$

$$i(t) = \frac{1}{R}\left\{v(t) - \frac{Q(t)l(t)}{\varepsilon A}\right\} \tag{1.3}$$

$$ml(t) = -\,bl(t) - k\{l(t) - l_0\} - \frac{Q^2(t)}{2\varepsilon A} \tag{1.4}$$

$$Q(t) = \frac{1}{R}\left\{v(t) - \frac{Q(t)l(t)}{\varepsilon A}\right\} \tag{1.5}$$

where, $v(t)$ = Controlling input voltage; $Q(t)$ = Charge in the capacitor; F = Coulomb force; C = Capacitance; $i(t)$ = Current through resistor; $l(t)$ = Air gap; l_0 = Initial gap (when $v(t)$ is zero); A = Plate area; ε = Permittivity in the air gap; R = Resistance in the circuit; m = Mass of the top plate; b = Damping constant; and k = Spring constant

By using lumped parameters the model can be represented pictorially (Fig. 1.9). The lumped parameters are pure and basic elements representing various forms of energies involved in the systems. The schematic model diagram of this microactuator is shown in Fig. 1.8. Bottom plate of the microactuator is held fixed and the top plate is free to move. Electrostatic actuation makes use of the Coulomb forces.

The last two equations are motion equations. By using these model equations the behaviour of the system can be analyzed and predicted prior to its real design. Varying the parameters such as input function and the physical variables such as mass, thickness, area etc. can achieve the desired response that we are after. For a specific output the set of design parameters can thus be calculated from the model equations. The formulation of model equation is simply called *modeling*. The method of calculation, observation, analysis, prediction and optimization using modeling formulae is known as *simulation*. For complex systems the simulation is performed by the use of computer-assisted tools. Computer assisted simulation tools contains all types of mathematical building blocks so that the designer can formulate any kind of model equations looking at the system and can subsequently analyze and predict the behavior. Even the designer can incorporate the physical properties (Bulk Modulus, permittivity, coefficient of resistance, etc.) of the material into the model equations. As an example, the Table A.5 in Appendix B shows some of the physical properties of polysilicon material, which may be used for FEM (Finite Element Method) analysis while simulation the MEMS structure to be made from that material.

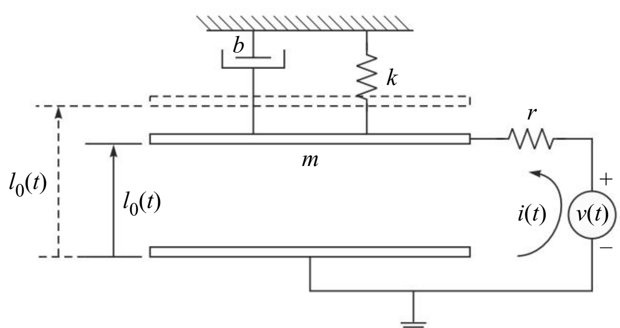


Fig. 1.9 *Model diagram of an electrostatic capacitive microactuator*

For some MEMS devices, it is tedious to derive analytic dynamic equations. Further, even if all the model equations are given, it is sometimes tedious to solve them analytically to predict the dynamic behavior of the system. Some reasonable approximations should be expected in the derived model.

1.19 MEMS PACKAGING AND DESIGN CONSIDERATIONS

Much like IC packaging, MEMS packages must have the ability to meet some important criteria, such as

(i) there should be good isolation between the non-sensing and sensing areas of the device,
(ii) there must not be any hindrance to the driving actions such as tilting, twisting, rotating, sliding, or vibrating,
(iii) efficient coupling at the link, junction, anchor area,
(iv) Unreliability issues due to the following reasons.

- Contamination
- Fusing
- Sticking
- Clamping
- Static overload
- Delamination
- Creep
- Fatigue

1.20 MICROINSTRUMENTATION

In order to achieve sophistication and availability miniaturization technology based instruments design concept is going to be adopted as advanced instrumentation platform for most of the scientific, industrial and academic studies. Microinstrumentation equipment is essentially useful where higher sensitivity, resolution, selectivity, fidelity and repeatability are desired. Miniaturization improves portability and speed of operation. Miniaturization can help the engineer to measure and analyze the physical, chemical and biological parameters of an application where space and weight are limiting factors. A typical microspectrometer can take less space while satisfying the required capabilities to measure, analyze and provide precise signals for further analysis and processing. Microinstruments can be applied in nuclear reactors, space shuttle, spectroscopy, surface analysis, tribology study, microtomography, imaging, and so on. Recently developed microinstrumentation equipment includes: micro positioning stage, microspectrometer, microgripper, surface plasom reonance immunator, microradar, etc.

1.21 SCOPE OF THE BOOK

The scope of MEMS technology is vast. It is considered as a multidisciplinary technology incorporating most of the fundamentals from a wide section of science and engineering. This book covers most of the topical subjects that the MEMS technology deals with Fig. 1.10 provides an abstract view of the MEMS discipline. The square blocks represent the subject areas and the semi-circular projections represent the interaction that exists among the subjects, emphasizing multidisciplinary scenario. The following enlisted fundamental topics have been covered.

- Micromachining
- Bulk and surface and micromachining
- Lumped-parameter based system description and modeling
- Mechanical, electrical and thermal properties of material
- Introduction to System-on-a-chip (SOC) and related mechanical and passive electronic systems (MES and PES)
- Mechanical MEMS
- Thermal sensors and actuators
- Micro Opto Electro Mechanical Systems (MOEMS)
- Magnetic sensors and actuators
- Radio Frequency MEMS
- Microfluidic devices (MFD) and systems
- Biomedical and chemical microsystems
- Introduction to nanotechnology and applications of carbon nanotubes (CNTs)
- Role of simulation and multiscale design approach
- Performance measuring parameters and device arrays

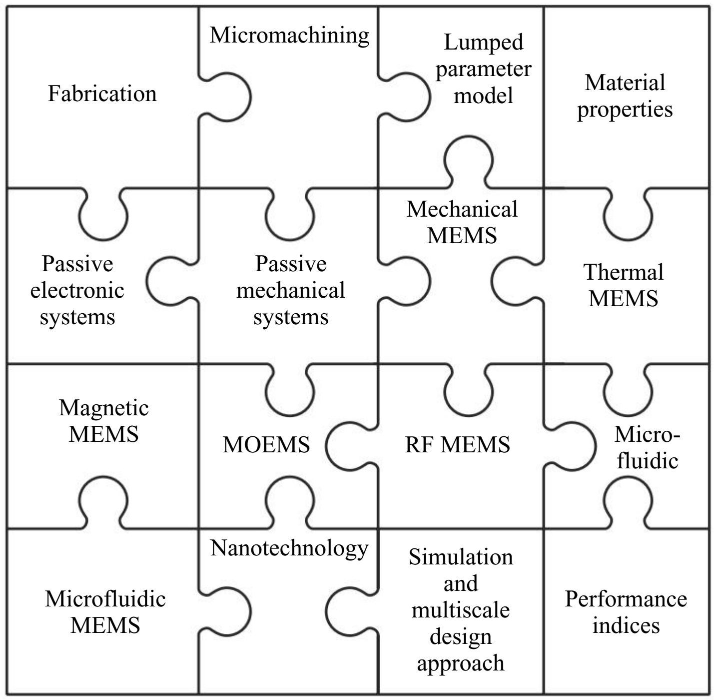


Fig. 1.10 *Scope of the book*

1.22 SUMMARY

This chapter is the introduction to Microelectromechanical Systems (MEMS). MEMS is an advanced technology that integrates complex electromechanical components and signal processing circuitry on a single wafer substrate similar to IC (Integrated Circuit) technology. These systems include both electrical and mechanical components. The major difference between the IC technology and MEMS technology is that, in the latter system the 3D mechanical components such as plate, beam, cantilever, diaphragm, gear, etc, as well as the electrical components such as inductors and capacitors, which were impossible to fabricate, are now fabricated by using dedicated design methods. MEMS is undoubtedly a class of sensor and actuator systems that are physically small, nominally in the order of micrometers. MEMS originally used modified IC fabrication techniques and materials to create these very small electromechanical components and devices. However, the advanced fabrication method significantly differs from the traditional modified technique and by virtue of that the new method had established itself as what is known as *micromachining*. Micromachining process makes stationary and/or moving 3D structures on a silicon wafer. Micromachining is categorized under two processes: Bulk micromachining and surface micromachining. The categorization is based on the way the 3D structures are created, i.e. whether from a bulk of material or from thin films.

The term micromachining has been derived from the conventional term machining which is defined as a process for the design of 3D structures and shapes for macro-level mechanical components and machine systems, such as crane, robots, engines, parts, and so on. Two important machining processes are micromilling and microdrilling. Since micromachining manufactures products whose dimension is in the order of micrometer it is justified to define the process as micromachining process.

At the beginning, the chapter reviews the scope of production engineering due to the reason that MEMS class of products are manufactured based on the concept of micromanufacturing. In fact, the scope of production engineering broadly includes three main sub-domains:

(i) Precision engineering and ultra-precision engineering,
(ii) Micromanufacturing, and
(iii) Nanotechnology.

Micromanufacturing, a sister discipline of Manufacturing, accommodates two important design methods such as Microelectronics and MEMS. A brief description in this respect is presented in this chapter.

MEMS classes of products are categorized as under:

(i) Mechanical MEMS,
(ii) Thermal MEMS,
(iii) Magnetic MEMS,
(iv) MOEMS,
(v) RF MEMS,
(vi) Microfluidic systems, and
(vii) BioMEMS.

Under all these categories two important devices such as microsensor and microactuators are common. This chapter briefly introduces the principle of operations of various types of sensors and actuators. Further, the chapter also highlights the application domains of each category. A road map to the following chapters is presented in the latter part of the chapter.

Points to Remember

- A technology that considers manufacturing of microscale, transducers, probes, capacitor, inductors, actuators, valves, gears, pumps, gyroscope, mirrors, switches, and so on, similar to semiconductor chips are referred to as microelectromechanical systems (MEMS).
- MEMS are highly complex, multiphysics and inter-interdisciplinary in nature entailing synergistic integration of many aspects of fundamental knowledge base, pursuance of unified understanding on related sub-domains is of paramount importance.
- NEMS is viewed as an interdisciplinary domain accommodating several disciplines, including general science! NEMS groups of products and systems fall under a well-known technology called *nanotechnology*.
- Mechatronics is defined as the synergistic integration of mechanical engineering with electronics and intelligent control algorithms in the design and manufacture of products.
- Manufacturing is the basis of production engineering.
- The precision engineering *is dedicated to the continual pursuit of the next decimal place.*
- Precision engineering includes design methodology, uncertainty analysis, calibration compensation, error compensation and controls.
- *Micromilling* and *microdrilling* are the two important processes.
- Micromilling suggests two different approaches: cutting tool based approach and focused ion beam (FIB) based approach.
- Microdrilling is characterized by drilling of ultrafine holes.
- Passive Electronic Systems are signal-conditioning circuits such as amplifier, ADC, filter, isolators, etc and Passive Mechanical Systems (PMS) are gear, crank, bearing, etc.
- PES, the microelectronic integrated circuits, can be thought of as the nervous system with sensors and actuators as the eyes and arms, respectively.

- The density of MEMS chip is measured in terms of number of components required to make up a particular device.
- Quite large numbers of sensors can be micromachined in a single platform as a *sensor fusion device* (SFD).
- The sensors are used for measuring physical parameters. The basic building block of a sensor is the transducer.
- The important transduction principles are thermoelectric, photoelectric, electromagnetic, magnetoelectric, thermoelastic, pyroelectric, and thermomagnetic.
- One good way to look at a sensor is to consider the important attributes such as stimulus, specifications, physical phenomena, conversion mechanism, material, response, ruggedness, stiffness, range, ability to measure physical parameters, and application field.
- An actuator is a control device that makes something move. The process of micromovement is called microactuation. Two ways of actuations such as translational and rotational are possible.
- The basis of actuation is the method of energy transduction. Principal energy transduction methods, called driving methods, are mechanical, thermal, electrical and magnetic method.
- The method of micromanufacturing of microelectronic devices is called *fabrication*. Fabrication is a sequential procedure. The important fabrication sequences are film growth, doping, lithography, etching, dicing and packaging.
- The term micromachining refers to the fabrication of 3D MEMS structures with the aid of advanced lithography followed by etching. Broadly, the micromachining falls into two categories such as Bulk micromachining and Surface micromachining.
- Mechanical MEMS mostly emphasizes on two classes of devices; mechanical structure based device and piezoelectric material based device.
- Thermal MEMS work on thermal phenomena. Thermal phenomena are basically described in terms of thermoelectric effect, Peltier effect, thermoresistivity, pyroelectricity and shape memory effect.
- Micro Opto Electromechanical Systems is abbreviated as MOEMS. MOEMS are MEMS but they handle optical signals.
- One very important combined process that is being used is the LIGA process. LIGA is a German acronym for lithography, electroplating, and molding.
- A Spatial Light Modulator (SLM) is a device that modulates light according to a fixed spatial pixel pattern.
- Soft ferromagnetic materials have found the most utility in microsensors, microactuators, and microsystems. Hard magnetic materials have several applications including storage devices.
- Magnetoresistive (MR) materials are used for sensing applications.
- Permalloy has good soft magnetic properties, high permeability, high magnetoresistive effect, low magnetostriction, stable high frequency operation, and excellent mechanical properties.
- RF MEMS is an emerging technology that plays a major role in accelerating the current growth in wireless communication. RF MEMS can be used for achieving (i) Transmission and reception, (ii) VCO tuning, (iii) RF band select filters, (iv) Intermediate Frequency (IF) filtering, (v) Time delay for phased-arrays, (vi) Variable Delay Lines (VDL), and (vii) Reconfigurable antennas design.
- The study of transportation of fluids and their mixtures at a microscale level is known as microfluidics.
- Microdevices used for analysis and detection of biomedical and industrial reagents are called bio- and chemodevices.
- The development of DNA sensors is considered as the most innovative molecular biology technology. The basic element of the DNA analysis system is the DNA probe(s). The process of DNA recognition is called hybridization.
- The hybridization can be achieved in many ways such as optical method, surface stress and electrical detection.
- The nanoscale devices and equipments can give added benefits in terms of further miniaturization, efficiency and resource consciousness.
- Carbon is suitable for the nanotechnology-based products.

- A fuel cell is an electrochemical system that generates electricity by converting chemical energy into electrical energy.
- MEMS devices are designed using micromachining technique. Prior to their design, it is desirable to study the behavior of the systems.
- The formulation of model equation is simply called *modeling*. By using lumped parameters the model can be represented pictorially.
- The method of calculation, observation, analysis, prediction and optimization using modeling formulae are known as *simulation*.
- Microinstrumentation equipment is essentially useful where higher sensitivity, resolution, selectivity, fidelity and repeatability are desired.
- Miniaturization improves portability and speed of operation.

Exercises

1. Name the scientist who proposed the possibility of manufacturing the ultraminiaturize system.
2. Expand the abbreviations MEMS and NEMS and define them.
3. Discuss the scope of production engineering discipline.
4. Distinguish between the terms mechatronics, micromechatronics and microinstrumentation.
5. Briefly outline the origin of precision engineering. Mention its application domains. What are the two important processes of precision engineering? With examples discuss the features and characteristics of micromilling and microdrilling process.
6. Enlist some of the applications of microdrilling process that you know.
7. Why is the acceptance of Integrated Circuit (IC) platforms more than the discrete component based solution? Name some IC chips available in the technology marketplace.
8. With suitable diagram discuss the multi-engineering design scenario of MEMS devices.
9. What do you mean by System-in-a-Chip (SOC) technology? Is it applicable to IC-based system? Suggest some examples, which could implement SOC concept.
10. What does the scale of integration mean? Draw the scale of integration plot of some of the important SOC devices. Plot the horizontal and vertical axes as the number of mechanical and electronic components, respectively.
11. What do you mean by next-generation MEMS? What additional features should a typical next generation MEMS have?
12. Briefly discuss the scope of MEMS discipline.
13. Write short notes on the following.
    (a) Microsensors
    (b) Microactuators
14. What important process sequences are followed in fabricating microelectronic devices?
15. Prepare a table to show some of the major wet enchants and their features.
16. Define machining and micromachining processes. How do they differ from each other? Define bulk micromachining and surface micromachining. Which micromachining process is better and why?

17. Write short notes on the following.
    - (a) Mechanical MEMS
    - (b) Thermal MEMS
    - (c) MOEMS
    - (d) Spatial Light Modulator
    - (e) Magnetic MEMS
    - (f) RF MEMS
    - (g) Microfluidic systems
    - (h) BioMEMS
    - (i) Chemo-devices
    - (j) Nanotechnology
18. Why is the study of modeling and simulation important for system design, particularly for MEMS design? What are the specifications of requirement as far as modeling is concerned? Give an example of how a parallel plate actuator system can be modeled. Represent the above system in pictorial form and describe each element.
19. What design considerations are involved in the packaging of MEMS systems?
20. Summarize the scope of MEMS topical subjects in pictorial form.

Chapter 2

Micromachining

Objectives

The objective of this chapter is to study the following.

- Machining and micromachining processes
- Lithgraphy methods such as UV photolithography, X-ray photolithography, electron-beam lithography, soft-lithography and their comparisons
- Various types of photoresist materials and their applicability
- Characteristics of thin film and deposition methods: LPCVD, PECVD, e-beam evaporation and sputtering
- Role of doping and doping methods (diffusion and ion-implantation)
- Various types of etching process. Dry etching and wet etching methods with examples of etchants and their comparison
- Clarification on isotropic and anisotropic etching
- Surface micromachining
- Bulk versus surface micromachining
- Surface micromachining process parameters
- Comparison, relative merits, and discussion on micromachining
- Wafer bonding method

2.1 INTRODUCTION

Machining is defined as the process of removing material from a work piece in the form of chips in order to obtain the exact shape and size of the work piece required. Machining is a well-defined process in 'machine design and tooling'. It includes several processing methods, which are usually divided into three main classes such as *cutting*—generally involving single-point or multipoint cutting tools, each with a clearly defined geometry; *abrasive processes*—such as grinding; and NTM (nontraditional machining)—utilizing electrical, chemical, and optimal sources of energy. Machining process handles macro components and machine systems. Micromachining bears the similar definition but the process handles microsystems. Micromachining has become a new technology, as there has been significant growth for the manufacture of components having dimensions less than 10 mm, upon which micro

features in the range of 1–500 μm are built. In particular, micromachining deals with microelectronic, micromechanical, micro-opto-electrical and micro-opto-mechanical systems. When these are collectively integrated in one platform then the system is known as MEMS systems.

This chapter describes fundamental of the micromachining process. Micromachining is classified under two different process technique such as bulk micromachining and surface micromachining. This chapter deals with both the techniques.

2.1.1 Micromachining

Micromachining is considered as a process as well as a technology that is utilized to structure wafer materials or thin films in order to fabricate miniature devices such as microsensors, microactuators and passive components (electronic amplifier, bridge circuits, post, gear, hinges, flexure, etc.) for microsystems functioning as electromechanical, optoelectronic and optomechanical systems.

Micromachining process manufactures microdevices in terms of designing or implementing the features in the bulk of materials such as silicon, quartz, SiC, GaAs, InP, Ge and glass. The features are implemented through several processes, which will be described in detail. However, it should be noted that one of the important sub-processes in the micromachining is *etching*. Etching is simply considered as removal of material. The etching is performed either on the substrate itself or on a preferred material layer deposited on the substrate. The very meaning of *substrate* is the material upon or within which a plant or animal lives or grows however; in the context of MEMS, instead of using the phrase 'plant and animal' we have to use 'MEMS components (also sometimes called features) and circuitry'. In essence, a substrate (also called wafer) is a bulk of base material without any interconnection pattern, over or within which discrete mechanical and electronic components and integrated circuits (IC) are built by micromachining process, which includes etching. Etching process is primarily used to remove a defined portion of the substrate in a particular manner so that the desirable shape can be obtained. Etching thus selectively removes material from the substrate. The specific material to be removed is determined by the etching solution, called *etchant*. Without loss of generality, Fig. 2.1 can provide an insight into the etching process. Some typical basic features (or structures) such as plates, steps, grooves, cantilever, diaphragm and post can be implemented through etching process.

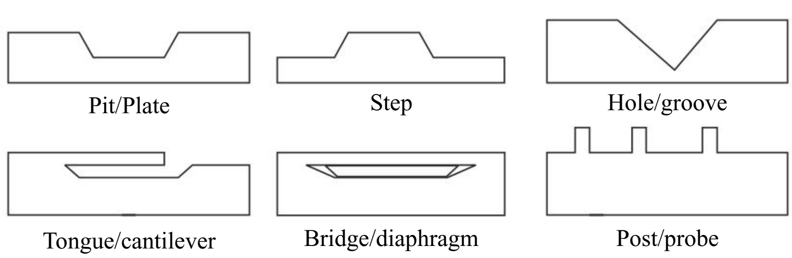


Fig. 2.1 *Typical basic MEMS structures can be created by etching the substrate*

Manufacturing of microdevices is referred to as fabrication. We will also use the word 'fabrication' as the synonym to micromachining throughout the book, but the term fabrication is more frequently used for IC (integrated circuit) manufacturing. As MEMS design approach is very compatible to the IC design method, we prefer to use them interchangeably for the same meaning.

If structuring is performed on wafer materials, the process is called bulk micromachining process and if it is performed on the thin film, the process is called surface micromachining process. There are considerable differences in the processing technologies of the two techniques mentioned above, leading

to differences in the fabricated structures. Comparison between the bulk and surface micro-manufacturing process is presented at the latter part of the chapter. However, a brief comparison can be found in Table 2.1.

Table 2.1 *Comparison between bulk and surface micromachining*

Property	*Bulk*	*Surface*
Processing complexity	0	0/+
Lateral dimensions	3–5 mm	100–500 μm
Vertical dimensions	100–500 μm	0.5–2 μm

Source: J. Micromechanics and Microengineering 8 (1998)

Both technologies are widely used and are being developed further. In some applications the technologies are in direct competition, whereas one dominates its counterpart. But they have many common approaches such as they heavily rely on the following methods, principles and processes.

- Wafer cleaning and deposition
- Photolithography and pattern transformation
- Doping—diffusion and ion implantation
- Etching
- Metallization (with Al, Au, Ti, Pt, Cr and Ni using sputtering, evaporation or electroplating)
- Packaging

In the sequel the above methods, principles and processes are described in more details.

2.2 PHOTOLITHOGRAPHY

It is already mentioned that the microcomponents/microfeatures are created either from the substrates or from the thin layer(s) of some specific material, say a silicon dioxide layer. Assume that we are interested in fabricating several simple microstructures such as *steps,* as shown in the Fig. 2.1. First, the oxide layer is deposited on the entire substrate. Note that the layer is not entirely used to create the desired components such as steps. This is due to the reason that a typical MEMS device might contain many components, which could be built from the same layer. Justifiably, several components are derived from the single layer(s). Therefore, the deposited layer has to be segmented and subsequently the selective portions be removed or etched away properly so that many desirable components (e.g. steps) can be created in respective segments. Depending on what type of components and what kind of MEMS structure is to be produced the layer at hand is segmented and subsequently etched. The entire process is usually achieved by a technique called *photolithography*. Photolithography is thus an essential as well as one of the starting sub-processes used to delineate the shape of MEMS structure in first place. Simply speaking, segmentation characterizes the removal of some portion from the layer, leaving the rest over the substrate. So the portions of the layer which need not be removed are to be protected by some means, and this is achieved by the use of a mask, which is prepared in advance. The foremost phase of the photolithography is therefore the preparation of a *photomask*. Once prepared, it is then transferred onto the layer.

Lithography is a Latin word which means stone-writing. Stone-writing is a process of printing that utilizes flat inked surfaces to create the printed images. Although, the principle is not utilized here in the context of MEMS design but the very meaning is similar. Photolithography is the process of using light

to create a pattern, i.e. mask, and subsequently transfer it onto the substrate (wafer). Photolithography is an optical means of transferring patterns into the substrate.

In the meanwhile, the substrate is chemically cleaned to remove particulate matter as well as any traces of organic, ionic, and metallic impurities on the surface. Sometimes it is necessary to pre-process the substrates by heating them in a chamber maintained at temperature ranging from 950–1100 °C (depending upon the material and applicability) in order to facilitate a uniform etch rate. The wafer cleaning process and photomask preparation can go alongside. Figure 2.2(a) shows a thin film of some selective material (say, silicon dioxide; depending upon the design requirement, however, other materials such as metals, alloys, etc. can also be used) deposited on a substrate of some other material (say, silicon). It is required that some of the silicon dioxide is to be selectively removed so that it only remains in particular areas on the substrate.

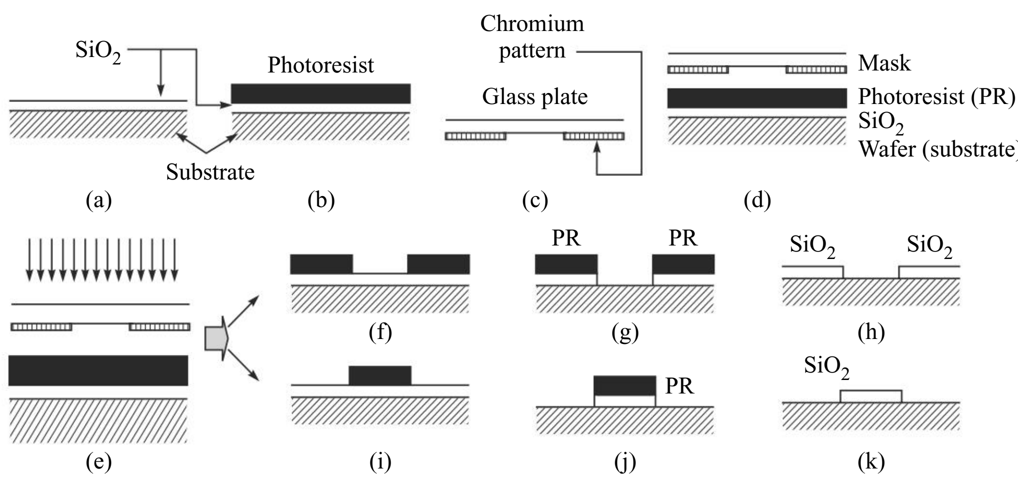


Fig. 2.2 *Photolithography; (a) A substrate with SiO_2 layer, (b) Additional photoresist layer (c) The mask—glass plate with chromium pattern, (d) The mask is on the top of the photoresist-SiO_2-wafer layers, (e) UV light falling on the photoresist, (f) After exposed to UV light (positive PR), (g) Third phase of photolithography in which the opening portion of the SiO_2 layer is removed (positive PR case), (h) Final phase of photolithography in which the PR is removed (positive photoresist), (i) After exposed to UV light (negative PR), (j) Third phase of photolithography in which the opening portion of the SiO_2 layer is removed (negative PR case), (k) Final phase of photolithography in which the PR is removed (negative PR)*

For this we need to produce a mask in first place. As mentioned, the photomask used in the photolithographic phase is a key component in the process. The mask is typically a glass plate that is transparent to ultraviolet (UV) light. The pattern of interest is generated on the glass by depositing a very thin layer of metal, usually chromium or gold. These masks are capable of producing very high quality images of micron and even sub-micron features (Fig. 2.2(c)). Typical width of the mask may vary from 3–50 µm. Photomasks are normally prepared by the help of computer-assisted software platform (Fig. 2.3).

The next phase of the process is the coating of a typical material on the wafer. The material should be sensitive to UV light. Usually organic polymers are chosen. This UV light sensitive polymer material is called photoresist. Once the photoresist coating is over, the photomask is placed over the coating as

shown in Fig. 2.2(d). Then the UV light is allowed to fall on the mask. When UV light falls on the photoresist (Fig. 2.2(e)) through the mask, a similar pattern is developed on the photoresist layer. This phenomenon is called mask transformation or pattern transformation, as a similar mask or pattern is developed on the imagable photoresist layer. The transformation is achieved by the combined effect of UV light and the composition of the photoresist. Photoresist can be either soluble or insoluble after being exposed to UV light (Fig. 2.2 (f) & (i)). Accordingly, there exist two types of photoresist; positive and negative photoresist, respectively. The soluble photoresist becomes weaker when exposed to UV light, on the other hand the opposite occurs to insoluble photoresist (negative type). In practice, the photoresist is washed away in the portion where the light was struck; conversely, the negative photoresist is not. This assures that resist which was not exposed to UV light is washed away forming a negative image of the mask. The process of washing away the material is called etching (more details later).

The photoresist is coated on the surface of the SiO_2 layer (in this typical example) by a process called *spin coating* which is described in Section 2.5.

In the third phase, the portion of the oxide layer that is now exposed through the openings of the photoresist is removed by some chemical processes (Fig. 2.2 (g) & (j)). Finally the photoresist is removed leaving the desired segmented oxide layer (Fig. 2.2 (h) & (k)). The result of Fig. 2.2 (f), (g) & (h) has undergone the process in which the positive photoresist is used and that of Fig. 2.2 (i), (j) & (k) undergone the process in which the negative photoresist is used. Figure 2.3 summarizes the photolithography process described above. Figure 2.2(h & k) shows steps, which are built using photolithography process.

Constraints on Exposing Energy UV light is known as the *exposing energy*. The source for UV light is a mercury arc lamp, called radiator, which has an output with spectral energy peaks at particular wavelengths. Depending upon the feature complexity, the thickness and the property photoresist

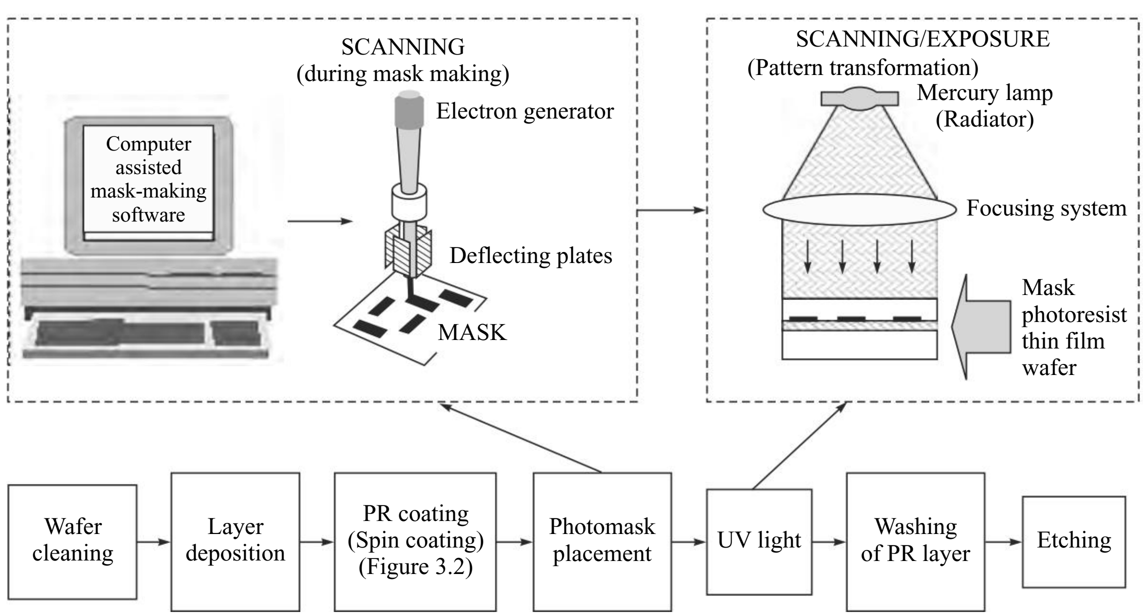


Fig. 2.3 *More details on photolithography*

materials, the exposing may be performed once or several times; accordingly we have single-exposure fabrication and multi-exposure fabrication process. The feature size of the bottom of the structure becomes smaller than that of its top after the development. This may cause the features to break near their bottom. Therefore, accurate dose of UV light is of paramount importance. Sometimes the additional exposure may cause large internal stress in the resist layer leading to a poor adhesion between the resist layer and the substrate. The internal stress may lift-off the structures from the substrate during the subsequent processes. Therefore, a compromise is sought. Light may fall on the entire area of the photoresist in one go or in sequence by employing scanning technique. Scanning is achieved by moving a small spot of light over the desired area of photoresist.

2.3 STRUCTURAL AND SACRIFICIAL MATERIALS

MEMS components and structures are manufactured through micromachining. The materials from which the microcomponents are built are called structural materials. The components and structures are micromachined in two ways; indirect and direct ways. When a microcomponent or structure is built from a previously micromachined micromold then the method is called indirect method (see soft lithography section). A mold is a 3D geometrical structure used for making its 3D image using other molten material. The process of shaping the material using a mold is called molding.

On the other hand, when the structures are created without using the molds, the micromachining process is called direct method. Whatever may be the ways, the structures are built from the selective materials called structural material. There exist several types of structural materials. Depending upon the application, e.g. whether the MEMS structure is for RF (radio frequency) application, biomedical application, fluidic application and so on, the structural materials are chosen. Some structural materials are SiO_2, SiN and Polysilicon.

Sacrificial layer is a layer, which is deposited on a specific region and later on is removed so that microstructures or features can be created. In Fig. 2.4, SiN is used as the structural material for

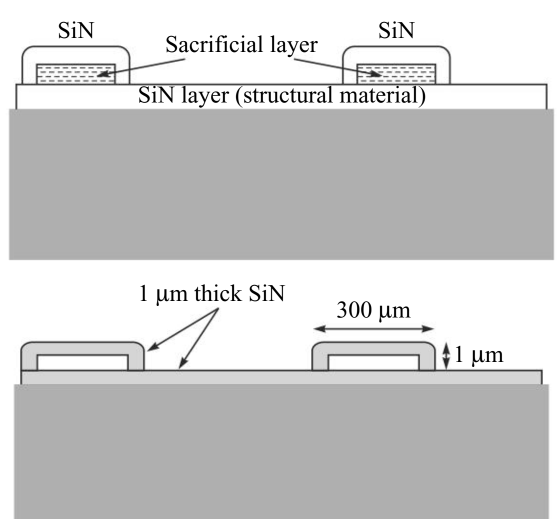


Fig. 2.4 *A microstructure showing the role of structural material and sacrificial material (a) Before etching (b) After etching.*

micromachining two bridge structures could be used as resonator in RF MEMS applications and Aerogel as the sacrificial material, which is etched away at a latter stage in order to obtain the freestanding bridges (resonators). Among sacrificial materials SU-8 and Aerogel inherit good physical, chemical and mechanical properties for which they are very popular. Although, the primary applications of SU-8 and Aerogel are to act as sacrificial material, they can also be used as structural material for micromolding applications.

Table 2.2 *Physical properties of the SU-8 photoresist*

Characteristic	*Value*
Absorption coefficient	2–40 cm^{-1}
Breakdown voltage	1.1 10^5V/m
Coefficient of thermal expansion	20–50 ppm/°K
Degradation temperature	~384.5 °C
Density	1.2g/cm^3
Film stress	19–16 Mpa
Glass temperature	5 °C ~250 °C
Loss tangent	0.08–0.14
Max sheer	0.009
Max stress	34 Mpa
Modulus of elasticity	4.02–4.95 GPa
Poisson ratio	0.22
Refractive index	1.575–1.8
Relative dielectric constant	3–45
Shrinkage	7.5%
Thermal conductivity	0.2 W/mK
Viscosity	0.6–15 pa.s

2.3.1 SU-8

SU-8 is an epoxy-resin negative photoresist widely used for fabricating MEMS structures. Figure 2.5 shows comb-like and gear-like structures, developed by the process of photolithography, in which SU-8 is used as structural material in Fig. 2.5(a) and as the sacrificial material in Fig. 2.5(b), respectively.

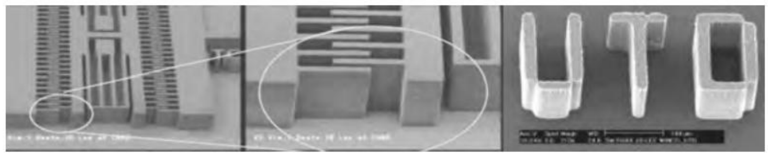
The advantage of using SU-8 photoresist is that high a*spect ratio* of approximately 20:1 3D microstructures can be built. Aspect ratio is a measure of geometry of 3D structure, quantified as a ratio of width to height. Inherited with excellent chemical and mechanical properties this polymerized resist has opened up wide number of novel applications in the field of MEMS. Structures fabricated with SU-8 resist can normally have a thickness of several hundred micrometers. Recently the SU-8 2000 series, which are formulated in cyclopentanone solvent have been introduced. These new series have improved coating and adhesion properties. In order to use SU-8 as the structural material it should be cured beforehand. Some important physical parameters of SU-8 photoresist is presented in Table 2.2.

2.3.2 Aerogel

Aerogel, a new class of material, invented in the 1930s but recently refined by NASA, can be used for fabrication of thin films in MEMS applications. Aerogels are highly porous nanostructured materials typically prepared by sol-gel processing. They possess the following unique characteristics.

- Lightweight
- High relative surface area
- Low dielectric constant
- Low thermal conductivity

Aerogel is virtually weightless solid measuring 0.00011 lbs per cubic inch. Note that thin air weighs about 0.00004 lbs per cubic inch. Chemically similar to glass, the mass of a micromirror can be reduced by 70–90%, increasing the resonant frequency and shock resistance, while preserving the optical surface area. Low density makes it useful as a lightweight structural material, and its super-high internal surface area makes it a super-insulating solid material. Further, aerogel has high etch selectivity for which it requires small etch-time as compared to other non-porous materials. Because of this property



(a)

(b)

Fig. 2.5 *(a) SEM photomicrographs of 20:1 aspect ratio (thickness 400 μm with minimum feature size of 20 μm) comb drive microstructures made using SU 8 resist (Source – U. of Texas at Dallas), (b) A MEMS gear mechanism*
Courtesy: MEMX Inc., California

aerogel is even used as sacrificial layer facilitating the release of large area structures. Table 2.3 shows some physical properties of silica and alumina aerogel films. Table 2.4 lists the etch-time of various etchants for 1 μm-thick silica and alumina aerogel thin films.

2.4 OTHER LITHOGRAPHY METHODS

The lithography method described in the previous section is called optical lithography because it uses UV light as the exposing energy. Other exposing energies such as X-ray and electron beam can also be drawn on. Besides the consideration of the source of exposing energy, the lithography is also characterized by hard- or soft-lithography based on the configurability of mask (described in this

Table 2.3 *Physical properties of aerogel films*

	Silicon aerogel	*Alumina aerogel*
Pore volume	1.16 cm^3/g	0.80 cm^3/g
Average pore diameter	72.0 Å	61.8 Å
Porosity	72.0%	70.8%

Source: Fan et al., 2001, *IEEE*

Table 2.4 *Etch time for 1 μm silica and alumina aerogel films in various Etchants*

Etchants	*1 μm silica aerogel*	*1 μm alumina aerogel*
49% HF	< 1 sec	< 5 sec
BOE[a]	5 sec	1 min
Piranha[b]	Non-etched	Etched
30 wt% KOH	< 5 sec	< 5 sec
AZ 400 K[c]	> 10 min	~1 min
XeF_2	< 5 pulses	No apparent etching (> 120 pulses)
RIE[d]	~2 min	No apparent etching (> 30 min)
DRIE	~3 min	> 10 min

[a] 40% Ammonium : 49% HF = 1 : 6
[b] Sulfuric Acid : Hydrogen Peroxide = 5 : 1
[c] DI water : AZ 400 K = 5 : 1
[d] Pressure ratio of reactant gases : (CF_4 : O_2 = 5 : 1), Power : (200 W)

Source: Fan et al., 2001, *IEEE*

section). Photolithography is a hard-lithography method. The last part of this section deals with soft-lithography method.

2.4.1 X-ray Lithography

X-ray lithography uses collimated X-rays as the exposing energy. Collimation is the act of restricting the size of the useful X-ray field to the region of interest. X-ray lithography is an easier, faster and very flexible lithography process. Sub-micron scale structures are fabricated with relatively good precision and structure quality compared to optical lithography. Further, X-ray lithography provides improved lateral resolution for which microstructures with relatively greater height as compared to optical lithography can be fabricated. Lateral resolution is associated with the complex non-vertical 3D geometries such as sloped sidewalls and curved surfaces. Tall structures in the range of some in the millimetre height with aspect ratio of the order of 100:1 are possible. In a single exposure, fabrication of large areas is achievable. Many optical MEMS structures (called MOEMS: MicroOptoElectroMechanical System) entails smoothed vertical sidewalls in order to minimize scattering loss[1]. In this respect X-ray lithography method is preferred. X-rays with short wavelength typically 0.1–1.2 nm make it possible to realize precise, nano-scale patterning permitting deep penetration into thick sensitive resists over large exposure areas. The X-ray exposing energy based method is also used to pattern a layer of substrate to form micromolds. Later on, the molds can be filled in by metals to make parts for other MEMS.

The interaction of matter with light and X-ray differs. As a consequence, compatible mask materials are chosen for X-ray lithography. Typical mask materials are diamond, gold, beryllium, polyimide, silicon or silicon carbide. The photoresist of choice is usually polymethyl methacrylate (PMMA), but SU-8 can also be used. SU-8 is more X-ray sensitive than PMMM. The copolymer resist made up of

1 The loss due to the effect of scattering. Scattering is a phenomenon in which the direction, frequency, or polarization of the light wave is changed because of discontinuities in the medium. Scattering occurs due to the roughness of a reflecting surface.

methyl methacrylate (MMA) and methacrylic acid (MAA) is approximately 10 times sensitive than that of PMMA.

Bending magnet beamlines with a bending magnet radius of 3 m and an X-ray scanner located approximately 10 m from the source point carries out X-ray exposing process. Figure 2.6 shows an X-ray lithography chamber. The chamber fielded with helium gas, has a mask holder and a resist stage. The photoresist is exposed through the mask as shown. The source of the X-ray is a synchronous radiator (SR). The radiator is attached to the storage ring. Storage ring is a container consisting of a set of magnets placed in toroidal manner around which charged particles from the SR followed by an accelerator is kept circulating until they are used. The ring can have typically 10 unit cells (magnet stations) with circumference of tens of meters. The X-ray is passed through the 200 µm of Beryllium windows and 50 µm of Kapton® polyimide film. Beryllium (used at the storage ring side) and Kapton (used at the chamber side) provide excellent way to isolate vacuum and other environments while allowing X-rays to pass through. Kapton® polyimide film is primarily used for insulation purpose. The wavelength of the ray can be between 0.1 and 1.2 nm. Typically the storage ring is operated at 1.3 GeV.

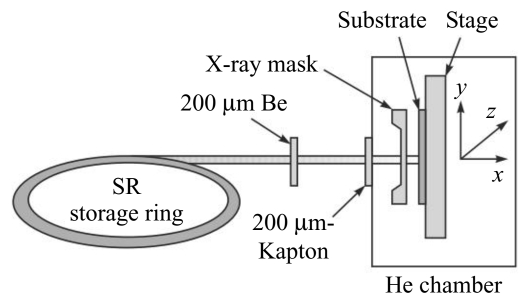


Fig. 2.6 *The X-ray lithography chamber. The chamber has a mask holder and scanning resist stage*
Source: Sugiyama et al., *J. Micromech. Microeng.*, 14, 2004

The calculation of absorbed X-ray energy distribution in the perpendicular cross-section of the resist (e.g. PMMA) is given as follows. The transmitted X-ray spectrum S_{TP} about an optional depth x_P through the membrane of the X-ray mask S_M is given by,

$$S_{TP} = S_M \times e^{-\mu_P x_P} \tag{2.1}$$

where μ_P is the absorption coefficient of photoresist. The X-ray energy absorbed by the photoresist is obtained by differentiation of the transmitted X-ray energy spectrum with respect to depth x, which can be generally expressed as:

$$S_{AP} = \frac{d}{dx} S_{TP} \tag{2.2}$$

The X-ray energy absorbed by the resist, about an optional depth x_p is

$$E_{AP} = S_{AP}\, d\lambda \times D \tag{2.3}$$

where, λ is the wavelength of the X-ray radiation and D is a quantity called exposure dose quantity.

2.4.2 Electron Beam Lithography

As usual the resist is patterned with the combined effect of the mask and exposing energy. The resist can also be patterned by electron beam. Hence the name of the method, electron-beam lithography, or simply e-beam lithography. E-beam lithography uses a focused electron beam to write ultrafine patterns. The beam can take the shape of a square or rectangle. E-beam method can produce structure with better resolution compared to other lithography method. The much shorter wavelength of the electrons is one of the reasons for achieving higher resolution microstructures. JBX-9300FS and IMPRIO™ are the lithography systems and can be used for micromachining of various wafer materials including simple and compound semiconductor materials with minimum resolution of 20 nm!

2.4.3 Soft Lithography using Polydimethylsiloxane

Soft lithography is another technique for the fabrication of micro- and nanostructures. It is a relatively recent technique that can also facilitate high-resolution patterning and molding of soft polymers materials, effectively. Implicating an efficient and inexpensive way of producing intricate microstructures, the soft lithography is used to microfabricate a wide range of materials, such as complex organic groups, sol-gel materials, colloidal materials, suspension solutions of salts and precursors, and carbon materials. Wide range of substrates including glass, plastics, ceramics, or carbon can be chosen. This new patterning technique developed at Harvard by Prof. George Whitesides in which a mask is made from an elastomeric material called polydimethylsiloxane (PDMS). Patterns with critical dimensions as small as 30 nm can be generated. Note that the photolithography uses a rigid photomask in the process. On the other hand, soft lithography, a non-photolithographic set of microfabrication method generate a patterned elastomer that can be used either as the mask, mold or stamp to develop micropatterns and microstructures. Photomasks are used to transfer patterns into thin films of photoresist using conventional photolithography. Molds are used to form supported and free-standing microstructures of various materials. Stamps are used in high-resolution lithography to print patterns of self-assembled monolayers (SAMs) on appropriate substrates. The stamps are usually prepared by casting prepolymers against *masters* patterned by other conventional lithographic techniques. The stamp provides ink that forms a SAM on a solid surface by a covalent chemical reaction. SAM is defined as the spontaneous organization of molecules to matter with geometric repeat symmetry at the molecular level. The transfer of a pattern from an elastomeric stamp to a solid substrate is based on conformal nanoscale interaction between substrate and stamp that allows transport of material from stamp to substrate.

PDMS block with patterned relief structures on its surface is the key to soft lithography and can be used advantageously in diverse processes requiring patterning. Polyurethanes, polyimides, and cross-linked Novolac™ resins (a phenol formaldehyde polymer) can also be used instead. However, PDMS have a unique combination of properties resulting from the presence of an inorganic siloxane and organic methyl groups. Some important desirable properties are given below. The PDMS are:

- Fluid at room temperature
- Can be converted into solid by cross-linking
- Provide a surface with low interfacial free energy (approx. 21.6 dyn/cm)
- Good chemical stability
- Molecules or polymers being patterned or molded do not react with the surface of PDMS
- Does not swell with humidity
- Passes gas easily
- Good thermal stability (approximately 186 ^{o}C in air)
- Optically transparent down to approximately 300 nm
- Durability (The PDMS block can be used more than 50 times over a period of several months)
- Stamps or molds can be deformed mechanically to manipulate the patterns and relief structures in their surfaces

The elastomeric mold is prepared by cast molding. The elastomer is poured over a *master*. The master is another mold, which is fabricated using microlithographic techniques such as photolithography or micromachining. The master has a relief structure on its surface. The elastomer is

then *cured* and *peeled* off. Usually a catalyst called curing agent such as a mixture of a platinum complex and copolymers of methylhydrosiloxane and dimethylsiloxane is mixed with the PDMS. Once mixed, poured over the master, and heated to elevated temperatures, the liquid mixture becomes a solid. Prior to this, the master is silanized by exposure to the vapor of $CF_3\ (CF_2)_6(CH_2)_2SiCl_3$ for about half an hour. Figure 2.7 describes the procedure for fabricating PDMS stamps. Some very specific applications of PDMS stamp/mask are

- Microcontact printing
- Replica molding
- Microtransfer molding
- Solvent-assisted micromolding
- Phase-shift photolithography
- Cast molding
- Embossing
- Injection molding

The most important application of soft lithography is to pattern gently curved surfaces. Table 2.5 shows the comparison between photolithography and soft lithography.

2.5 THIN FILM DEPOSITION

MEMS device fabrication processes involve the deposition of thin films onto the surface of a wafer. Thin films are developed for variety of purposes. For instance, the films can form microstructure, conducting, and insulating or even semiconducting layers. Film deposition can be achieved using spin coating, oxidation and chemical vapor deposition (CVD). There exist a distinguishing feature between the terms *layer* and *thin film*. The thin film can be understood as a coating or layer that is so thin that only its surface properties are used for exploitation. On the other hand, the thickness of a layer is relatively high. Since this distinction is not very stringent, in many cases 'layer' and 'thin film' are used

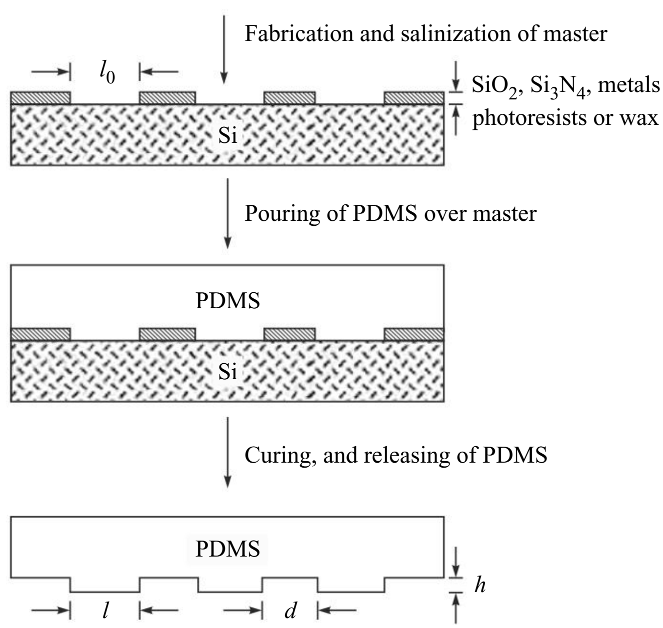


Fig. 2.7 *Procedure for fabricating PDMS stamps from a master with relief structures on its surface (Xia and Whitesides, Annu. Rev. Mater. Sci. 1998)*

Table 2.5 *Comparison between photolithography and soft lithography*

	Photolithography	*Soft lithography*
Definition of patterns	Rigid photomask (Patterned Cr supported on a quartz plate)	Elastomeric stamp or mold (a PDMS block patterned with relief features)
Materials that can be patterned directly	Photoresists (polymers with photosensitive additives)	Photoresist[a,e]
	SAMs on Au and SiO_2 and SiO_2	SAMs on Au, Cu, Ag, GaAs, Al, Pd, and SiO_2[a]
		Unsterilized polymers[b-e] (epoxy, PU, PMMA, ABS, CA, PS, PE, PVC)
		Precursor polymers[c,d] (to carbon and ceramics)
		Polymer beads[d]
		Conducting polymers[d]
		Colloidal materials[a,d]
		Sol-gel materials[c,d]
		Organic and inorganic salts[d]
		Biological macromolecules[d]
Surfaces and structures that can be patterned	Planar surfaces 2-D structures	Both planar and nonplanar Both 2-D and 3-D structures
Current limits to resolution	~250 nm (projection) ~100 nm (laboratory)	~30 nm[a,b], ~60 nm[e], ~1µm[d,e] (laboratory)
Minimum feature size	~100 nm (?)	10 (?) – 100 nm

[a-e] Made by (a) µCP, (b) REM, (c) µTM, (d) MIMIC, (e) SAMIM. PU: polyurethane: PMMA:poly(methyl methacrylate): ABS:poly(acrylonitrile-butadiene-styrene): CA: cellulose acetate: PS: polystyrene: PE: polyethylene: and PVC: poly(vinyl chloride)

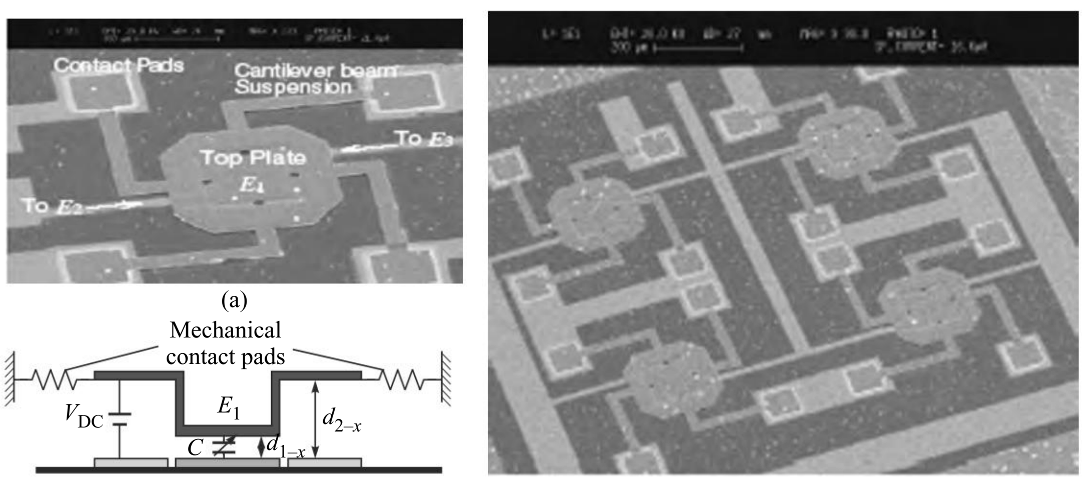
Source: *Annu. Rev. Mater. Sci.* 1998.

loosely and in certain calculation the bulk as well as surface properties are taken into account and exploited. Sometimes thin films are referred to as diaphragms. Literally the diaphragm, which is a component in MEMS structure, should be treated as a 2 D thin film rather than a 3 D object.

Depending upon the applicability, materials for the thin film are selected. The commonly used materials for developing thin films are silicon, silicon nitride, silicon dioxide, germanium, gallium arsenide, polymers, metals, ceramics or diamond. The mechanical properties that draw attention while fabricating the thin films include residual stress, elastic modulus, yield strength, ultimate strength, creep and wear. Primarily thin films are developed as transducer element for sensory applications, microphone diaphragm for instance. The applicability may be extended to develop dielectric layer, plate of the parallel plate capacitor (PPC), electrode, connectors, resistors, inductors, ferromagnetic storage films, plated through-holes and multi-layer interconnects. Alumina, beryllium oxide, aluminium nitride, fused silica, ferrites, titanates, glass, and plastic are usually used as substrate materials on which thin films are developed.

2.5.1 An Example

Figure 2.8 shows an application of thin films in MEMS. The figure shows a typical variable capacitor, which may be used for tuning applications (See Chapter 9). It is a three-parallel-plate tunable capacitor. Of the three parallel films, the capacitor has one suspended top plate and two other fixed bottom plates. The top plate and one of the two fixed bottom plates constitute the electrodes of the variable capacitor, whereas the other fixed bottom plate and the top plate are used as an actuator. The movement of the actuating plates is achieved by electrostatic method. As the plates move, the separation distance changes causing a change in capacitance. Variable capacitors are called varactors. Air or piezoelectric materials are placed in between the electrodes. Varactors are used to select (tune) the desired frequency components. In order to achieve wide range of tuning capacitor gangs are fabricated. Figure 2.8(c) shows a micromachined wide-range varactor.



(b) (c)

Fig. 2.8 *(a) A thin film based MEMS varactor for tuning applications, (b) Its schematic representation showing mechanical contact pad, electrostatic actuation voltage, (c) A wide range tunable varactor composed of 4 capacitors*
Source: Zou et al., *IEEE* Copyright, 2000.

2.5.2 Thin Film Development Processes

Thin films are deposited by employing the following methods.

- Spin coating
- Thermal oxidation
- LPCVD (Low Pressure Chemical Vapor Deposition)
- PECVD (Plasma Enhanced Chemical Vapor Deposition)
- E-beam evaporation
- Sputtering

Spin Coating In spin coating, as the name implies, the wafer is rigidly placed on a rotating base called vacuum chuck. The photoresist polymer solution is put on the solid wafer. The solvent in the solution is usually volatile. The rotor is then allowed to rotate at high speed (Fig. 2.9). Rotation causes the solution to spread over the wafer surface while the volatile solvent evaporates. Depending upon the required thickness and uniformity the rotation is continued. When the solution is spun off the edges of

the substrate due to centrifugal force the rotation is stopped. The following formula is useful while calculating the time of spin t (seconds), for a given rotational speed w (radians per second).

$$t = \frac{3\mu}{4\rho w^2}\left(\frac{1}{h^2} - \frac{1}{h_0^2}\right) \tag{2.4}$$

From the above equation the thickness can also be obtained, which is given by,

$$h = \frac{h_0}{\sqrt{1 + 4Kh_0^2\, t}} \tag{2.5}$$

where, K is caked system constant given by $K = \rho w^2/3\mu$, h_o is the initial height, h is the final height or thickness (meter), μ and ρ are the viscosity (poise) and density of the solution, respectively.

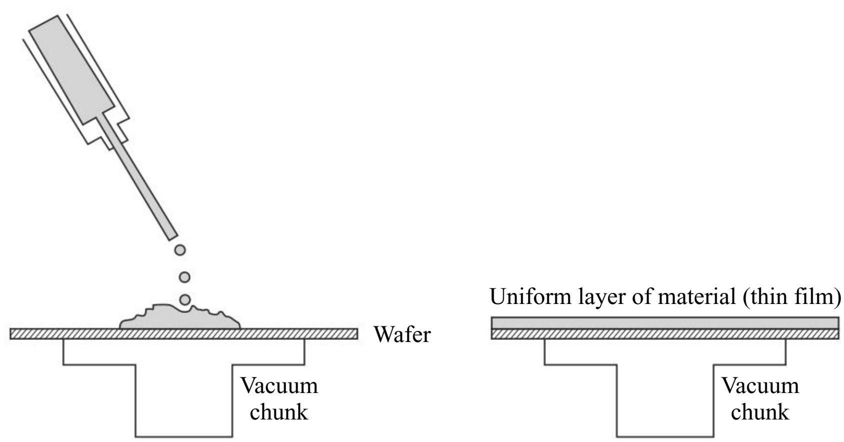


Fig. 2.9 *Spin coating, (a) Before spinning (b) After spinning (MEMS and Nanotechnology Exchange)*

Example Find out the desired viscosity to density ratio of a photoresist solution to be prepared for the spin coating by using a vacuum chunk. The allowed time of rotation of the chunk within a production process is 2 minutes and the desired height of photoresist is 50 microns. The rotor rotates at 10,000 rpm. Assume that the original height is 500 microns.

Solution: From Eq. 2.4 the viscosity to density ratio can be found as:

$$\frac{\mu}{\rho} = \frac{4tw^2 h_0^2 h^2}{3(h_0^2 - h^2)}$$

Given: Required time of rotation, t = 2 minutes = 120 seconds

Speed of rotation, w = 6000 rpm = $2\pi \times 6000/60$ rad./s = 200π rad./s

Initial height (before spinning), h_0 = 500 micron = 500×10^{-6} meter

Final desired height, h = 50 micron = 50×10^{-6} meter

Putting the above values directly in the above equation, the viscosity to density ratio can be obtained as:

$$\frac{\mu}{\rho} = \nu = \frac{4 \times 120 \times (200\pi)^2 \times (500 \times 10^{-6})^2 \times (50 \times 10^{-6})^2}{3\{(500 \times 10^{-6})^2 - (50 \times 10^{-6})^2\}} = 0.1595\ m^2/s$$

The viscosity to density ratio, ν is known as the kinematic viscosity.

Thermal Oxidation Oxide films having thicknesses ranging from 60 to 10000 angstroms is easily grown through this process. The oxidation occurs in a diffusion furnace which typically consists of heating unit, temperature measurement and control unit and process chamber where the wafers undergo oxidation. Thermal oxidation is classified either as *dry oxidation* or *wet oxidation* depending upon which oxidant species is used. Typically, O_2 and H_2O can become oxidant for dry and wet oxidation. Silicon dioxide (SiO_2) is one of the important materials in developing thin films. The formation of SiO_2 on the surface of a silicon wafer is accomplished in the diffusion furnace. The silicon substrate is exposed to an oxidizing environment of O_2 or H_2O which is called oxidant. In order to expedite the growth rate of oxide layer the oxidation is carried out at extremely high temperatures usually between 700–1300 °C. In this example, the reactions for dry and wet oxidation are governed by the following equations.

Dry oxidation: $Si \text{ (solid)} + O_2 \text{ (vapor)} \rightarrow SiO_2 \text{ (solid)}$

Wet oxidation: $Si \text{ (solid)} + H_2O \text{ (vapor)} \rightarrow SiO_2 \text{ (solid)} + 2H_2 \text{ (vapor)}$

LPCVD Chemical vapor deposition or CVD is a generic name for a group of processes that involve deposition of solid material from a gaseous phase. LPCVD (Low Pressure Chemical Vapor Deposition) is a technique in which one or more gaseous forms of materials are used to form either a film (could be an insulating or conducting) on another surface such as a wafer under low pressure and relatively high temperature conditions. LPCVD creates thin films of material on a substrate via chemical reactions. In an example, Tungsten (W) films of thickness 50–450 nm get deposited in the temperature range 450–600 °C and pressure range 0.5–4.5 Torr. Thin film with compressive low stress values are obtained at 0.5 Torr while moderate tensile values are obtained at about 4.5 Torr. Resistivity up to 75 μΩcm is obtained at 600 °C. In another typical example, to achieve resistivity of the order of 2–3 μΩcm, a conducting film of volatile copper dipivalylmethanate material may be developed by using LPCVD. The LPCVD process can also deposit material on both sides of the substrate at the same time.

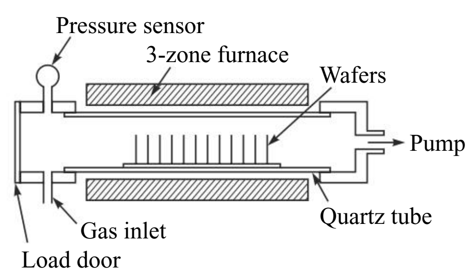


Fig. 2.10 *Schematic diagram of an LPCVD system*

LPCVD is performed in a reactor at temperatures up to 900 °C. The reactor is a tubular chamber made up of quartz. The chamber is divided into three zones, each corresponds to different temperature range (Refer to Fig. 2.10). The wafers are placed on the bottom of the quartz tube and the chamber is heated through the three-zone furnace. The deposited film is a product of a chemical reaction between the source gases supplied from a source into the reactor through the gas inlet. Source gases are from the material, which need to be deposited. A feedback pressure sensor is fitted to monitor and regulate the pressure within the chamber. This unit is called backing pump. Excess pressure must be pumped out through the outlet!

LPCVD based polysilicon films exhibit tensile or compressive residual stresses that depend on the deposition temperature. The temperature and pressure maintained through the feedback system within the furnace therefore, govern the residual stress of the deposited film. By suitably fixing the temperature of the chamber, structures made from the polysilicon films can be designed to display deliberate curvatures, because of stress, during deposition resulting in a desired optical device, such as focusing mirror. Although, the residual stress is almost unwanted, it is sometimes exploited within the design process itself, as in case of curved mirror. The film stresses also vary depending upon the deposition

pressure. If the films are aged at certain temperature for about several hundred hours the film can possess excellent mechanical stability. The main problems with the process are the high deposition temperature (sometimes higher than 800° C) and the relatively slow deposition rate. Table 2.6 shows some characteristics of thin films developed from various materials.

For designing MEMS devices operating at cryogenic temperatures using LPCVD method, values of the Young's modulus and fracture strength are considered as the design parameters. Silicon nitride thin films are preferably chosen for such low temperature applications. The Young's modulus of LPCVD silicon nitride thin films varies from 260.5 GPa at 298 K to 266.6 GPa at 30 K, and the fracture strength ranges from 6.9 GPa at 298 K to 7.9 GPa at 30 K.

Table 2.6 *Properties and characteristics of some thin film materials (Data obtained from MicroFAB BREMEN GmbH)*

Material	*Thickness range*	*Thickness non-uniformity*	*Deposition temp*	*Refractive index*	*Mechanical stress*	*Dielectric constant*
silicon dioxide	50 to 1500 nm	$< \pm 10\%$ (across wafer)	450°C	1.46 ± 0.05		
silicon nitride	20 to 150 nm	$< \pm 2\%$ (across wafer)	770°C	2.0 ± 0.05	1 GPa (tensile)	
silicon oxinitride	50 to 500 nm	$< \pm 5\%$ (across wafer)	800°C	1.8 ± 0.05	760 MPa (tensile)	4.8
silicon oxinitride (low stress)	50 to 500 nm	$< \pm 5\%$ (across wafer)	800°C	1.58 ± 0.05	230 MPa (tensile)	4.1
standard polysilicon	50 to 1500 nm	$< \pm 5\%$ (across wafer)	620°C	3.9 ± 0.1	200 MPa (tensile)	
tensile stress polysilicon	100 to 1500 nm	$< \pm 5\%$ (across wafer)	<620°C	2.0 ± 0.05	300 MPa (tensile)	

PECVD Plasma Enhanced Chemical Vapor Deposition (PECVD) is an important deposition method for the fabrication of thin films in VLSI and MEMS. Much like LPCVD, PECVD is used to form a solid film on the surface of a wafer by the use of one or more gaseous reactors (materials). Exclusively, in this case the operating temperature is much lower than the LPCVD process. Traditional CVD process creates thin films of material on a substrate via the use of chemical reactions. On the other hand, PECVD causes the reactive gases to decompose via the electrical discharge, thereby generating plasma. The deposition is enhanced by the use of a vapor containing electrically charged particles or plasma. Plasma is a fluid of positive ions and electrons in a quasi-neutral (quasi means having some resemblance) electrical state.

If it is required to deposit two thin films one above the other, bottom layer is preferably deposited by the process of LPCVD and the second layer by using PECVD at lower temperatures since after the first layer has been deposited LPCVD can no longer be used for the deposition of the second layer because the process will melt the layer previously deposited. Therefore, any subsequent deposition should be performed at low temperatures, with PECVD. PECVD has many advantages including low thermal budget requirement and development of flexible film properties. PECVD makes it possible to tailor film properties in order to obtain specific device characteristics. Figure 2.11(a) illustrates a schematic diagram of a typical PECVD system. Essentially the PECVD system has the following functional units.

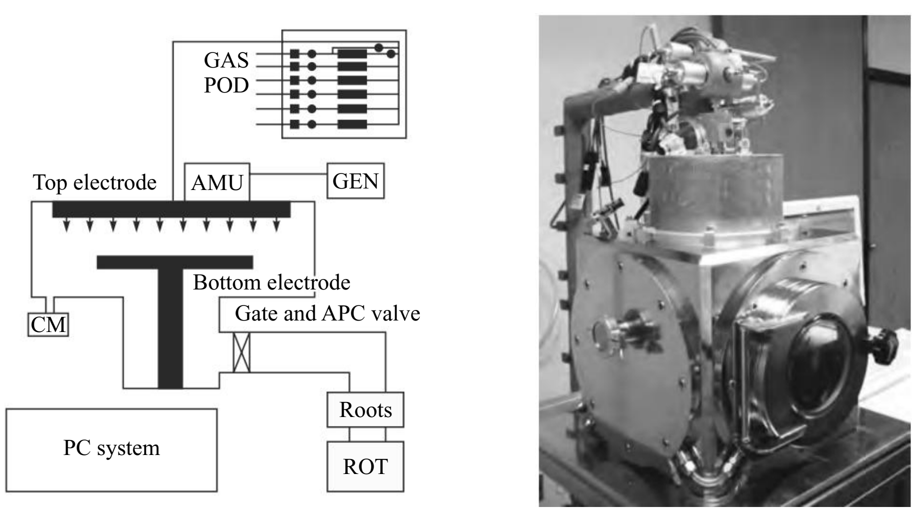
- Controlled gas lines such as gas pod with Mass Flow Controller (MFC)
- Electrode with high uniformity showerhead
- Gas inlet to top electrode with Generator and Automatch Unit (AMU)
- Bottom substrate electrode which is heated and grounded
- System control unit with PC platform
- Backing pump with Automatic Pressure Control (APC) valve
- Pressure gauge (called Capacitance Manometer; CM)

PECVD process begins with a wafer placed on a heated bottom electrode of a parallel plate configuration. A reactive gas from the gas pod is allowed to enter into the chamber that provides the elements from which the film is to be deposited. If it is required to develop either silicon, germanium or tin nitride film, the corresponding homoleptic dimethylamido complexes such as $M(NMe_2)_4$ are to be fed; (where, M = Si, Ge, Sn; Me = CH_3).

The plasma is generated within the chamber by a special arrangement. The plasma, which is now available between the electrodes, bombards the reactive gases causing further dissociation.

The electrical discharge phenomenon is a kind of ignition of the plasma that heats up the tool components and the wafer. Since ultimately the process is considered thermally driven, any deviation from thermal equilibrium will disrupt the deposition. In order to maintain thermal equilibrium, means such as backing pump, gate and APC and pressure gauge are equipped. Figure 2.11(b) shows the photograph of NPE-4000, a PECVD system from Nano-Master Inc. NPE-4000 PECVD system is capable of depositing high quality films on longer substrate size. It is equipped with a stand-alone PC-based control system that uses RF shower-head electrode of Hollow Cathode RF plasma source with Fractal Gas Distribution to generate plasma.

In order to verify the thickness of the film being deposited, real-time measurement and analysis systems are inbuilt into the PECVD systems. *In situ* single point two-color laser interferometer is commonly used to monitor the thickness of thin transparent films in real-time. The instantaneous



(a) (b)

Fig. 2.11 *(a) Schematic diagram of PECVD system (Courtesy – OXFORD Instruments); (b) Photograph of NPE-4000, a PECVD*
Courtesy: Nano-Master Inc.

change of film thickness is calculated by comparing the measured laser reflection interference to that calculated by a *priori* model. The deposition rates of the film can also be determined. It takes about 1–2 seconds to calculate and subsequently display the thickness and the deposition rate of the thin film, on-line. The use of two-color laser interferometry improves the accuracy of the growth rate measurement.

The PECVD process operates at lower temperatures down to 300 $^\circ$C. This lower temperature capability is due to addition of plasma energy to the CVD environment in the form of a glow discharge and is obviously the effective substitution of the thermal energy encountered in LPCVD. The extra energy is supplied to the gas molecules by the plasma. However, in this case the quality of the films tends to be inferior since the processes runs at low temperatures. Further, unlike LPCVD most PECVD deposition systems can only deposit the material on one side of the wafers.

E-beam Evaporation Electron beam evaporators are used to deposit thin films onto the substrates and are faster compared to other depositing techniques. Evaporation is achieved by locally heating a source material above its vapor pressure with an electron beam. A filament surrounding the evaporant rod is operated at high current of approximately 5 amperes to emit electrons. Evaporation is done under a high vacuum chamber. The source is an evaporant rod and behaves as a terminal, which is supplied with a positive high voltage of about 2.5 kV.

While in operation the evaporant rod attracts the electrons of the filament and gets heated. This in turn makes it possible to evaporate the source material. Due to very high electron density on the tip of the rod, high temperatures can be achieved. As the source material evaporates, it forms a thin film on the samples (substrates). By controlling the filament current the temperature can be controlled and hence the rate of evaporation of the source material, so that a wide range of film thickness from sub-monolayer to multi nanometers can be produced. Eventually, the film is produced from the tip of the evaporant. Three units such as power supply, source controller and the sweep controller control the electron beam. The power supply controls the current; the source controller is used to vary the voltage evaporant terminal, while the sweep controller controls the tip position when very small and spotted films are desired to be deposited. Sometimes a crystal oscillator is used to determine rates of evaporation and the resulting film thickness. E-beam evaporation method is very fast with better deposition control. Thick films and multiple coatings can also be achieved through this process. Figure 2.12 shows a multi-pocket e-beam evaporator.

Fig. 2.12 *E-beam evaporator*
Courtesy: tectra GmbH

Sputtering Sputtering is a process by which atoms are extricated from the surface of a target source material as a result of collision with high-energy atom or ion particles, generated from a plasma set up. The phenomenon is exploited for developing a thin film onto a substrate. The target source can be a metal, semiconductor or insulator. Upon impact (Figs. 2.13(a) and (b)), the high-energy particle tears away the source material, which retains the chemical and physical properties of that material after the sputtering is over. The extrication occurs as a result of the momentum transfer from the bombarding ion particle. Unlike other vapor deposition techniques, there is no melting of the material.

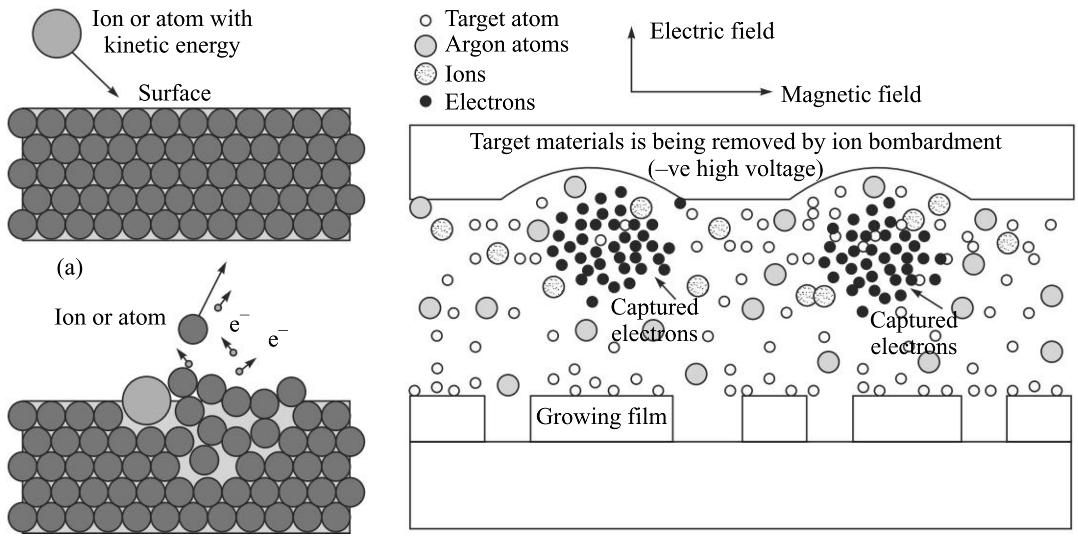


Fig. 2.13 *Sputtering mechanism (a) Before the impact, (b) After the impact, (c) An illustration of film formation because of sputtering effect*
Courtesy: Thermo Vacuum Generators

In contrast to CVD technique, sputtering based deposition method is known as Physical Vapor Deposition (PVD) technique. It is a physical process that takes place in a vacuum chamber in which the target material and substrate are essentially placed at the top and bottom, respectively. The evacuated chamber makes it possible to settle down the tore material onto the surface of the substrate, eventually depositing a thin film (Fig. 3.13(c)). Ionized argon can be used as the bombarding particle. The efficiency of sputtering primarily depends on three major factors such as the physical and mechanical properties of the source material, the mass and the energy of the bombarding particles. The thin film deposition sequence is as follows.

- By employing physical means the material to be deposited is first converted into vapor.
- The vapor is then transported across a region of low pressure from its source to the substrate.
- The vapor undergoes condensation on the substrate to form the thin film.

Sputtering has been one of the most widely used techniques for depositing even various metallic films on wafers, including aluminium, aluminium alloys, platinum, gold, TiW, tungsten, etc. The thickness of the coating is proportional to the applied voltage or equivalently the discharge current that is responsible for generating plasma and hence subsequently the high-energy atoms or ions. Determining the operating parameters and adjusting the deposition time can control the thickness of the film, which may vary from a few nanometers up to several micrometers. Sputtering has additional benefits over control of the alloy composition when compared with e-beam evaporation technique.

There are two ways of accomplishing sputtering mechanism; diode sputtering and magnetron sputtering. Traditional diode sputtering utilizes parallel plates, which are powered by potential difference of several kilovolts. The top negative cathode plate serves as the source material and is bombarded by ions from the plasma setup. This method is also called DC diode sputtering technique. Typical values of chamber pressure, temperature and deposition rate in a diode sputter are 0.5 to 10 Pa, 100 to 300 ^{o}C and

0.02 to 0.2 µm/min, respectively. The features of diode sputtering technique is that the process is simple, allows relatively easy deposition, operates at high deposition pressure and has relatively high substrate temperature. This method has disadvantages in terms of slow deposition rates, high-voltage and current requirements and severe charging-arching possibilities.

During the process, electrons escape from the target and move around the chamber randomly, causing radiation and other phenomenological effects including heating of the target. These electrons are called wasted electrons. It is very essential to confine the electrons in a definite region of the chamber. The region, called plasma glow region, is in the vicinity of the cathode. Magnetron sputtering technology overcomes the wasted electron problem.

The most widely used sputter technology for film deposition is based on the magnetron sputtering. Magnetron sputtering system takes the advantages of magnetic field by using magnets, which are placed behind the target material (cathode). These magnets capture the escaping or wasted electrons and confine them to the immediate vicinity of the target. Fundamentally, magnetron is defined as an electron tube constituted by the interaction of electrons with the orthogonal steady electric and magnetic fields. The interaction causes the escaped electrons to oscillate within the plasma glow region. Had they not been treated by the use of additional magnetic field, they would have interfered unnecessarily. Thus the use of magnetic field can enhance the sputtering rate, hence the deposition rate at lower pressure. The deposition rate is approximately 10 times faster than the diode sputtering method. Magnetron sputtering is particularly suitable for non-magnetic metals and alloys. Sputtering is also possible on inorganic materials. Table 2.7 shows the relative sputtering rates of some material with respect to copper.

There are several advantages of sputtering based film deposition. These are,

- Large-size targets are common
- Simple deposition of thin films with uniform thickness over large wafers
- Control over the alloy composition
- Control of film properties
- Cleaning of the substrate in vacuum prior to film deposition is possible avoiding additional cleaning process
- Unlike e-beam evaporation, the target and surface of the wafer do not get heated during the process

Table 2.7 *Relative sputtering rates of some material*

Material	*Etch rate w.r.t. Cu*	*Material*	*Etch rate w.r.t. Cu*
Ag	2.16	SiO_2	0.45
Au	1.76	Ta	0.43
Cu	**1.00**	Si	0.39
Al	0.73	W	0.39
Zr	0.65	Ti	0.38
Ni	0.65	Mg	0.26
Cr	0.60	Al_2O_3	0.15
Mo	0.53	C	0.05

Courtesy: Thermo Vacuum Generators.

2.6 IMPURITY DOPING

The process of adding impure material to a material of interest is called *doping* and the impure materials involved in this process are called *dopants*. In order to understand the utility of doping, we like to give an example of a microelectronic device such as a diode. The semiconducting diode is also known as junction diode since it is fabricated from two materials such as n-type or p-type. A pure semiconductor (e.g. Ge, Si or GaAs) is called *intrinsic* semiconductor and an impure one is called *extrinsic* semiconductor. Both n-type and p-type semiconductors are extrinsic semiconductors. Extrinsic semiconductors are obtained from the intrinsic semiconductors by the process of doping. The impurity material, called dopant, is deliberately added into the pure semiconductor in a controllable manner. Dopants are of two types, *donors* and *acceptors*. In the context of semiconductor technology, while the donor materials must have five valence electrons (Pentavalent) an acceptor atom must have three valence electrons (Trivalent)[2]. Doping is an important process, which determines the concentration of the excess electrons or holes. If donors are added, the intrinsic semiconductor becomes *n-type* semiconductor and if acceptors are added it becomes *p-type* semiconductor. n-type semiconductor materials have loosely-attached *free-electrons* and p-type semiconductor materials have loosely-attached *free-holes*. The electrons and holes, in the respective extrinsic semiconductors, are called the charge carriers. When a potential difference across the n-type or p-type material is established, the loosely-attached excess electrons or the hole will flow to the appropriate terminal. That is, in n-type the electrons will move from negative terminal to the positive terminal of the applied voltage and in p-type the holes will flow from +ve terminal to the –ve terminal of the applied voltage. In a semiconductor the current is contributed by both electrons and holes. In n-type material electrons constitute *majority* charge carriers, while in p-type materials holes constitute majority charge carriers. p stands for positive, signifying that a hole is a positive charge carrier, which is simply opposite to that of electron. The difference in concentration of dopant in a semiconducting material significantly influences the operational characteristics of the device. So the doping process provides interesting and useful properties as far as design, development and, indeed, fabrication of electronic devices are concerned.

Doping is also an essential process in MEMS designs. Note that many MEMS devices also include electronic devices and signal handling circuitry (called passive systems – see Chapter 4) such as filter, amplifier, data converter, isolator, and so on. Moreover, some components of the MEMS devices (e.g. transducer filament of a sensors device) may require doping in order to obtain the desired electrical properties. Doping process can be used to produce various films through variable conductivity. For instance, impurity doping modifies the resistivity of polysilicon. Boron and phosphorous are the most commonly used dopants in polysilicon. Figure 2.14 illustrates typical doping profiles. The doping concentration and resistivity is plotted as a function of the diffusion depth. Doping also facilitates controlling the etching process (described later in this chapter). There are two types of doping:

- Diffusion based doping
- Ion implantation based doping

2.6.1 Diffusion

A calculated amount of specific dopants are carefully diffused into the desired material in a controlled manner to achieve the required properties and behavior. Diffusion occurs if there is a concentration gradient. Diffusion process affords a relatively well-controlled method of introducing dopants precisely

2 For more information on bonding refer any book on solid state devices from Tata McGraw-Hill.

at shallow depths. Diffusion occurs when evaporated dopants are allowed to pass on the surfaces of the sample material (could be a wafer), that are placed in a chamber containing inert gas. The sample material is kept at an elevated temperature for several hours. The evaporated dopants are then allowed to pass on the surfaces at a controlled rate. In practice, the dopants are carried by the inert gas. The inert gas is actually bubbled through a liquid compound of the dopant that are then allowed to pass on the surfaces of the sample material which are placed in a chamber at high temperature. Evaporation based diffusion is a dynamic process since the inert gas is allowed to flow at some controlled rate. In some process the dopants are simply placed on the top of the material by painting, spinning, or other technique. The evaporation method is popular as this approach is more reproducible and cleaner.

2.6.2 Ion Implantation

Ion implantation refers to a doping process whereby dopant ions are introduced into the material of interest to change its property. Ion implantation is used as an alternative to dopant diffusion. Implant

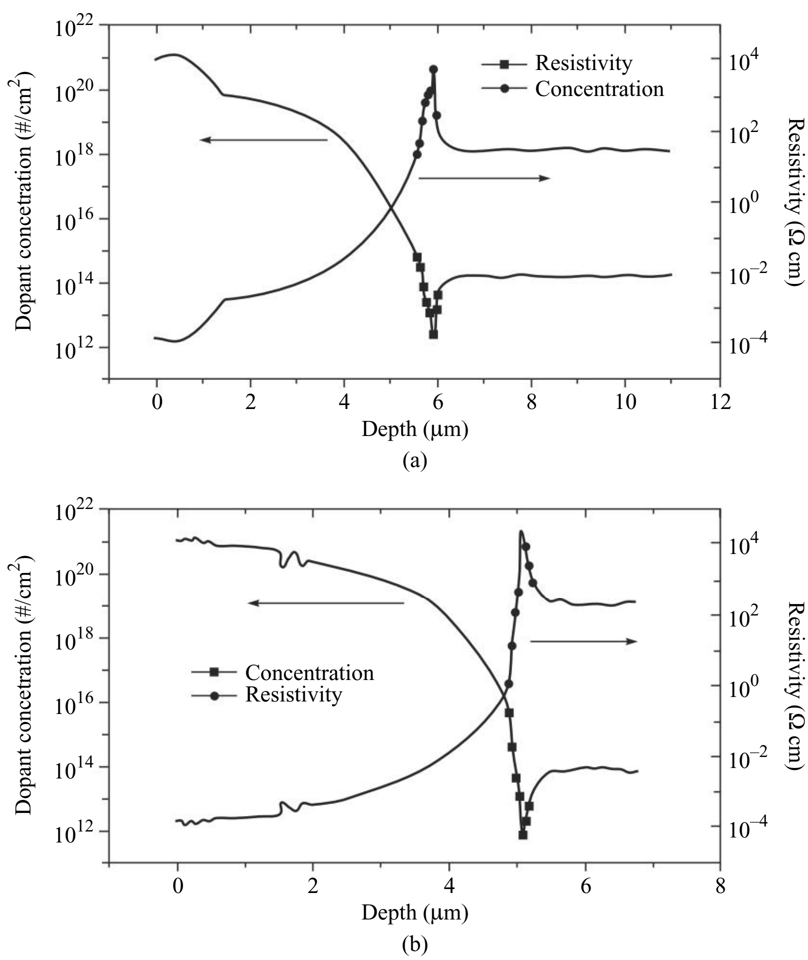


Fig. 2.14 *Doping profiles and resistivity of doped region. (a) Boron doping. (b) Phosphorus doping*
Source: J. MEMS, 8/4 Dec. 1999

affords a relatively simple means of placing a known number of atoms in a geometrical material. Accelerating the dopant ions in a strong electric field and directing them towards the surface of the material of interest achieves doping. The dopant ions are accelerated by several kilovolts and therefore possess enough energy to penetrate a solid surface at room temperature. The whole unit is housed inside a large vacuum chamber. The important criteria considered while implanting the dopants is the quantity of dopant (called dose), correct area, i.e. geometrical configuration, depth of implantation and the efficiency of implantation. The implanted dopants should be electrically active and there should be no damage to the implanted dopants. Ion implantation is a lower temperature process than diffusion. The sequence of the process follows (Refer Fig. 2.15).

- Source of the required dopant (solid, liquid or gas) is placed appropriately.
- The dopant is then ionized.
- The ions are accelerated through an electric field to acquire sufficient energy (usually in the range 1 keV to 1 MeV (1 eV = 1.60219×10^{-19} Joules)). Typically, this range of energies can facilitate average depths of implant from 0–10 μm.
- The analyzer magnet then filters out unwanted ions.
- The high-energy ions are directed towards the target material through the horizontal and vertical scanning unit.
- The horizontal and vertical scanning unit has electrodes which deflect the ion beam electrostatically and allow it to scan across the material so that the ions can be implanted at a desired area.
- After colliding with the material the ions lose their energy and eventually come to rest inside the target sample material.

The acceleration voltage determines the depth of penetration. The ions travel a distance, called the range, which is much larger than the penetration distance. The penetration is called the *projected range,* which depends on the energy of the incident ion. The dose can be monitored and controlled by calculating the ion beam current. Not all the ions are implanted perfectly. Some of the high-energy ions displace some of the atoms of the target before coming to rest. This disturbs the overall expected properties of the target. In order to nullify their adverse action, the ion implantation process is always followed by another process called *annealing*. Annealing is a process of heating a material at suitable temperature and then cooling it at a suitable rate, in order to restore the desired properties.

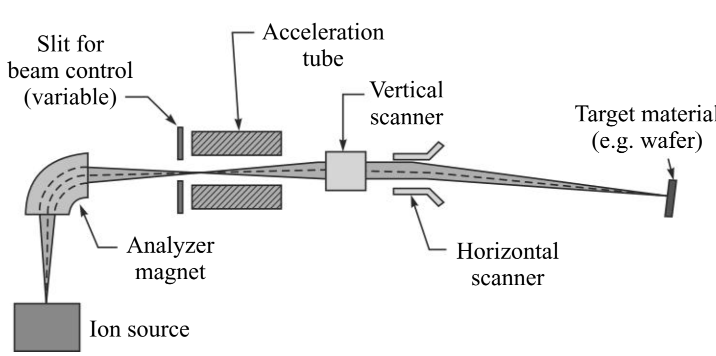


Fig. 2.15 *Schematic diagram of ion implantation based doping*
Courtesy: Microelectronics Process and Device Simulation Center, U. of Glasgow

2.7 ETCHING

Etching is a process, which makes it possible to selectively remove the deposited films or parts of the substrate in order to prepare a desired patterns, shapes, features, or structures.

It is necessary to etch the thin films (previously deposited) on a substrate or the substrate itself, in order to form a functional MEMS structure. As opposed to chemical vapor deposition, etching removes material from the deposited layer. The portion of SiO_2 layers and subsequently the portion of photoresist layers shown in the Fig. 2.2(f) and (i) and Fig. 2.2 (g) and (j) require etching in order to obtain the structure shown in the Fig. 2.2 (h) and (k), respectively. Some portion of the silicon substrate may be required to etch to form features shown in Fig. 2.1.

Two classes of etching process are common. They are *wet etching* and *dry etching*. Wet etching removes the material selectively through chemical reaction. The material is immersed in a chemical solution, which reacts and subsequently dissolves the portion of the material, which is in contact with the solution. Materials not covered by the masks are etched away by the chemical solutions while those covered by the masks are left undissolved. On the other hand, dry etching sputter the material using reactive ions or a vapor etchant. Each etching process is further classified as shown in Fig. 2.16. Figure 2.17 shows some of the photographs of MEMS structures after etching.

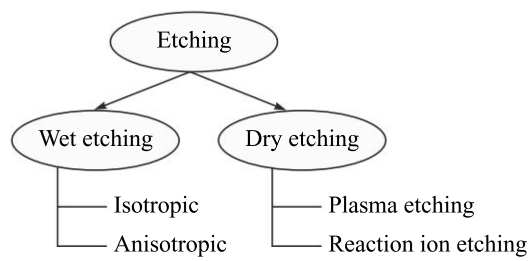


Fig. 2.16 *Classification of etching process*

2.7.1 Wet Etching

Wet etching is the simplest method. It is important to note that the liquid solvent should not change the chemical properties of the dissolved material such as photoresist, SiO_2, etc. Sometimes the wet etching process involves more than one chemical reaction and is, in fact, applicable only to multi-layer structures which require sequential etching.

The etchant is usually a mixture of acidic solutions. The selectivity of the etchant plays a major role in wet etching. Selectivity refers to how effectively the etchant reacts with the material without corrugating other materials deposited in the structure for other purposes (anchor, posts, etc.). An etchant may attack the mask or the substrate itself making significant changes in the etching profile, which is not desired. So appropriate combination of etchant, mask material and substrate are to be selected prior to fabrication in order to obtain good results. The sequence of operations within the wet etching process fall under three sub-activities:

- Diffusion of the etchant to the surface for removal. The operation is carried out at room temperature or slightly above, but preferably below 50°C.
- Establishment of reaction between the etchant and the material being removed.
- Diffusion of the reaction byproducts from the reacted surface. This activity can also be called cleaning.

The dissolution of material due to chemical reaction may not be uniform in all directions. This characteristic of etching is called directionality. Etch directionality is thus a measure of relative etch rates in different directions usually vertical versus lateral. The wet etching rate is crystallographic plane

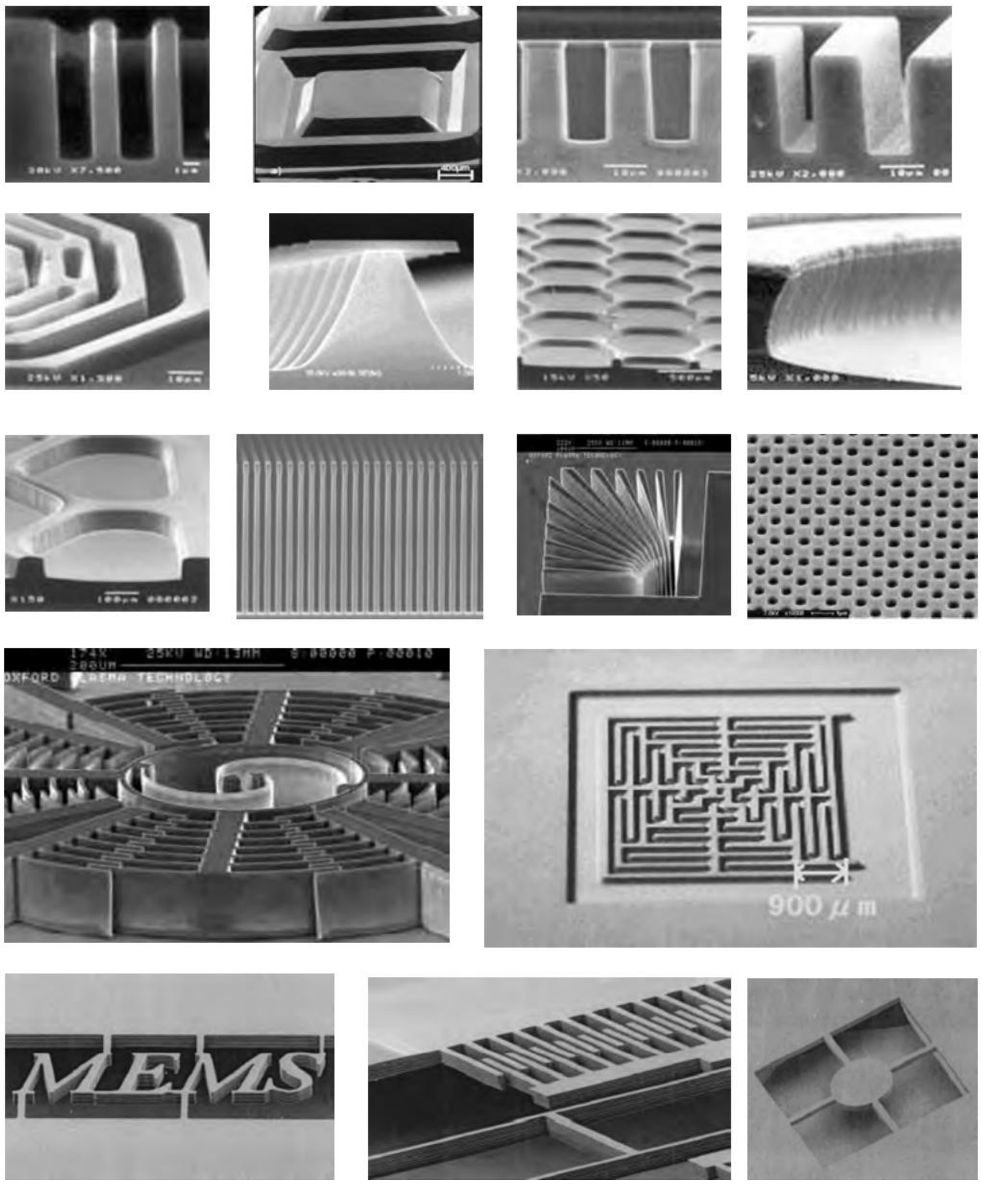


Fig. 2.17 *Some of the MEMS structure after etching. The dimension is in the order of 100–200 micrometers*
Courtesy: Samco Inc. and Sintobrator Ltd.

orientation dependent. That is the dissolution depends on the crystal structure of the material in question and, in fact, materials are characterized by whether they are anisotropic or isotropic in nature. Accordingly, the wet etching can be classified as whether the process etches the material anisotropically or isotropically. Anisotropic materials will dissolve in the way shown in Figs. 2.18(a) and (b), that is, the etch rates are not same in all directions (Etching rate of various planes of Silicon and SiO_2 at various temperatures are presented in Table App-7, in Appendix B). Anisotropic etching is considerably a highly

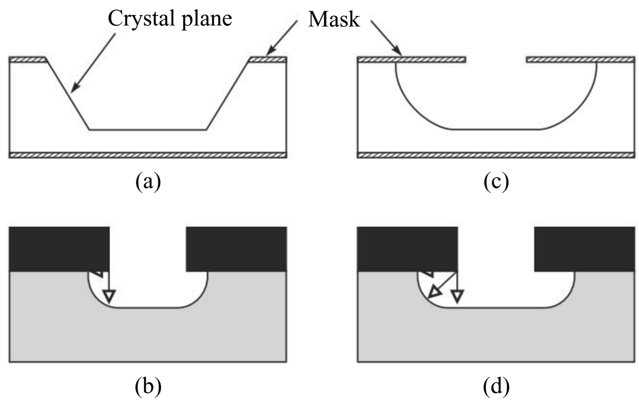


Fig. 2.18 *(a) Anisotropic, (b) Anisotropic showing directionality, (c) Isotropic, (d) Isotropic showing directionality*

directional etching process with different etch rates in different directions. The crystallographic planes of the material that etch slowly are called stop planes. The isotropic material will dissolve uniformly in all directions as shown in Figs. 2.17(c) and (d). The name isotopic comes from the word *iso* which means same. In isotopic etching materials are removed uniformly from all directions and it is independent of the plane of orientation of the crystal lattice. A comparison of wet and dry etching is given in a tabular form in Appendix B.

Note that the MEMS structure may require anisotropic or isotropic etching. Rectangular (or square) cavities, hemispheres, cylinders and mesas are achievable on the (100) substrate through wet anisotropic etching. Anisotropic etching is useful in producing Vee grooves, pyramids, and channels into the surface of the wafer. Given a specific material the etchant can also be selected to etch the material either isotropically or anisotropically. If it is required to fabricate holes of conical structure, i.e. wider towards the top and narrower toward the bottom on the crystalline silicon, then anisotropic etching will be obvious choice and we have to look for anisotropic etchants like alkaline chemicals.

The most commonly used isotropic etchants for silicon etchings are mixtures of hydrofluoric acids (HF), nitric acid (HNO_3) and acetic acid (CH_3COOH). HF: HNO_3 :CH_3COOH or H_2O is called HNA. The ratio of hydrofluoric acids, nitric acid and acetic acid is 3:2:5, by volume. Nitric acid consumes the Si surface to form a layer of SiO_2, which is dissolved by the HF. The final reaction can be expressed as follows.

$$Si + HNO_3 + 6\ HF \rightarrow H_2SiF_6 + HNO_2 + H_2 + H_2O$$

HNA etchant is an autocatalytic solution. The HNO_3 drives the oxidation of the silicon, while fluoride ions from HF then form the soluble silicon compound H_2SiF_6. Water can be used as diluent instead of acetic acid. But CH_3COOH is preferred as it prevents the dissociation of the nitric acid and preserves the oxidizing power of nitric acid, which is depended on the undissociated nitric acid. Controlling its reaction rate is more complicated than just fixing the ratios of the reagent components.

A wide variety of alkaline solution used for anisotropic etching of silicon include KOH (Potassium Hydroxide), EDP (Ethylenediamine Pyrocatechol), N_2H_4-H_2O (Hydrazine), NaOH, LiOH, CsOH, NH_4OH (Ammonium hydroxide) and ammonium hydroxides with addition of alcohol. Alkaline organics such as ethylenediamine, choline (trimethyl-2-hydroxyethyl ammonium hydroxide) or hydrazine with additives such as pyrocathechol and pyrazine are also used. Of the above, the anisotropic etching of (100)-oriented silicon wafers in an aqueous KOH is considered as a standard etching process in micromachining.

Similarly SiO_2 can also be etched through wet etching process. The reaction is as follows.

$$SiO_2 + 6\ HF \rightarrow H_2 + SiF_6 + 2\ H_2O$$

In practice, water-diluted HF with ammonium fluoride (NH_4F) is used as etchant formulation for SiO_2. Table 2.8 describes some of the materials and compatible etchants for wet etching.

Table 2.8 *Material versus etchant*

Materials	*Etchants*	*Remarks* (*Selective: Can etch slowly*)
Silicon Dioxide	Hydrofluoric acids (49% in water); Hydrofluoric acids NH_4F:HF (6:1); Buffered Hydrofluoric acids or Buffered Oxide Etch (BOE)	• Selective to Silicon • Film density and doping parameters determines the etch rate • Dilution of Hydrofluoric acids and buffered HF reduce the etch rate
Silicon nitride	Hydrofluoric acids (49%) H_3PO_4:H_2O (boiling temperature 150°C)	• Etch depends strongly on film density, Oxygen and Hydrogen • Selective to Silicon dioxide
Aluminium	H_3PO_4:H_2O:HNO_3:CH_3COOH (16:2:1:1)	• Selective to Si, SiO_2, and photoresist
Polysilicon	HNO_3:H_2O:HF(+CH_3COOH) (50:20:1)	• Etch rate depends on etchant composition
Single crystal Silicon	HNO_3:H_2O:HF(+CH_3COOH) (50:20:1) KOH:H_2O:IPA (49 wt. % KOH)	• Etch rate depends on etchant composition • Selective based on crystallography
Titanium	NH_4OH:H_2O_2:H_2O (1:1:5)	• Selective to $TiSi_2$
Titanium nitride	NH_4OH:H_2O_2:H_2O (1:1:5)	• Selective to $TiSi_2$
Photoresist	H_2SO_4:H_2O_2 (125°C)	• For wafers without metal

Source: Prof. Tayo Akinwande, MIT

In case of anisotropic etching, the corners of the structure do not etch uniformly resulting in a loss of the desired structure. For instance, convex corners are laterally undercut. This is due to the reason that some crystal planes etch faster than the others, as already mentioned. The adverse effect of corner etching can lead to either *overcut* or *undercut*. The effects are unwanted where a symmetry and precision 90° corner is essential. An undercut ratio is defined as the ratio of undercut to etch depth. When sharp rectangular corners are desired, it is necessary to take preventative measures prior to etching by employing some additional process. The precaution measure is known as corner compensation. The corner compensation is considered to be mandatory in order to achieve the precision desired geometry. Consider the overcut situation. One way of achieving the corner compensation is by adding extra structures at those corners susceptible to overcut so that after etching the desired accurate convex corners are obtained. The compensating material is actually added prior to the mask transformation process. Another way of achieving the corner compensation is to modify the mask layout itself without adding the compensation material. In this case, it is essential that the mask layout at the corner portions be modified in order to make it effective. The amount of mask modification and the shape depend on the

type of etchants and shape of the corner. By using saturated KOH solutions with isopropanol (IPA), the shortcomings can also be reduced. The depth of etch depends on temperature and etchant concentration. Undercutting although undesirable for 90^o rectangular corners, but useful for machining structures like convex corners and suspended membrane.

Besides directionality described above the etching process can also be described by other important figure-of-merit parameters, which are expressed in the Table 2.9.

Table 2. 9 *Figure of merit parameters*

Etch rate (R)	Rate of removal can be 0.1 micron/min. Important issues are throughput and control
Etch uniformity (U) $\frac{R_{high} - R_{low}}{R_{high} + R_{low}}$ R_{high} = maximum etch rate R_{low} = minimum etch rate	Expressed in percentage change in etch rate in the following manner. • Across a wafer • From wafer to wafer • Within a lot • From lot to lot
Selectivity $S_{fm} = \frac{R_f}{R_m}$; $S_{fs} = \frac{R_f}{R_s}$; R_f = film etch rate; R_m = mask etch rate R_s = substrate etch rate	Two terms are defined: S_{fm} : Film to mask selectivity S_{fs} : Film to substrate selectivity Usually they are the ratio of the etch rate of various materials such as the film to photoresist or the film to substrate
Anisotropy (A) $\frac{R_{vertical} - R_{lateral}}{R_{vertical}}$ $R_{vertical}$ = vertical etch rate $R_{lateral}$ = lateral etch rate	Measure of directionality of the etch. A = 1 implies perfect anisotropic etch A = 0 implies to isotropic etch
Undercut	Ability or measure of the lateral extent of the etch per side
Substrate damage	The damage occurred on the substrate due to chemical attack, displacement damage or abrasive damage. It could be physical and/or chemical damage.

Source: Professor Tayo Akinwande, MIT

2.7.2 Etching Control

When the desired shape and depth of the structure is reached further etching is not necessary so etching should be stopped. As an instance, while fabricating thin membrane sufficient precision measures are taken to control the etching. In order to control and stop etching during the process, means are evident. There are many techniques, by which the etch rate can be controlled and monitored. The etching process can be controlled through etchant agent, etchant composition, total amount of material etched, etchant temperature and diffusion effects. Light can also affect the etch rate. Some of the important

etch stop methods are boron diffusion etch stop, electrochemical etch stop and SOI (Silicon on Insulator) etch stop. Boron diffusion method is based on the fact that anisotropic etchants do not attack boron-doped (p+) silicon layers (Fig. 2.19). Material, from which the MEMS structures will be obtained, doped with both boron and germanium, for example, can be etched much slower than undoped silicon. Boron p+ is diffused from the back of the substrate before the anisotropic etching is performed. Electrochemical technique uses lightly doped p-n junction (being the wafer) as cathode and a counter electrode as anode in the etchant KOH. The bias voltage is applied between the wafer and anode. The wafer is immersed in the solution. In this process a p-type substrate can be etched and stopped at the p-n junction. Another etch-stop technique is employed in terms of a change in composition of material. As an example a layer of silicon dioxide between two layers of silicon can facilitate etching control due to the reason that many etchants of silicon do not react with silicon dioxide. The method is based on SOI etch stop.

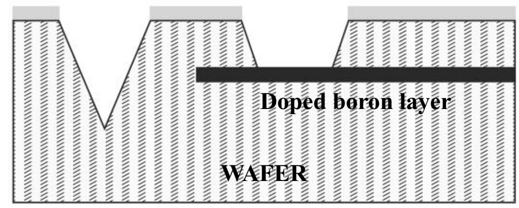


Fig. 2.19 *Boron diffusion can stop etching*

Wet etching has its own disadvantages as listed below.

- Limited resolution
- Risks due to the chemical exposure
- Problems related to adhesion to the substrate
- Formation of bubbles during etching process degrading the etch rate
- Sometimes incomplete and non-uniform etching
- Etchants are costly

2.7.3 Dry Etching

Despite its simplicity in controlling the etching in bulk micromachining, wet etching is not a flexible and reliable process. A dry etching process is an alternative choice. Dry etching does not utilize any liquid chemicals or etchants to remove materials. This etching process is primarily used in surface micromachining process (described later). The main advantages of dry etching are that the process eliminates handling of dangerous acids and solvents, uses small amounts of chemicals, both isotropic or anisotropic etch profiles are possible and good process control. The etching is accomplished by any of the following methods.

- By dry chemical reactions that consume the material
- By momentum transfer in terms of physical removal of the material (similar to sputtering)

Dry chemical etching is also known as plasma etching. The process sequences of dry plasma etching are given below:

- Reactive species generation in a plasma setup
- Species diffusion to the surface of the material subjected to etching
- Species adsorption on the material surface
- Chemical reactions between the species and the material takes place
- Formation of volatile byproducts
- Desorption of the byproducts from the surface
- Diffusion of the desorbed byproducts into the bulk of the gas

The gas plasmas are typically generated in a vacuum chamber at room temperature. The chamber is pumped to a pre-defined pressure. The process gas is introduced. A radio frequency (RF) electromagnetic (EM) field is applied to the electrodes in the chamber producing glow-discharge plasma. Within the plasma various gaseous species such as ions, free radicals, electrons, photons, neutrals, and ozone are produced. These species make highly active, low temperature plasma that can etch material selectively and quickly.

On the other hand, the physical sputtering is known as ion beam milling. Physical sputtered etching is based on the microscopic reaction steps as follows.

- Electrons are accelerated by RF (radio frequency) or microwave electric fields.
- They are then allowed to collide with suitable precursor molecules to produce ions, atoms, and radicals in order to produce a complex mixture of reactive species.
- These reactive species strike the surfaces that are in contact with them. Adsorption of these species on the surface takes place in terms of chemical reactions between the species and the material being etched.
- After striking on the surface the complex species are transferred to a volatile byproducts.
- Finally the desorbed byproducts are diffused into the bulk of the gas.

Both etching processes are widely used in micromachining. Both processes look similar, but the distinguishing feature is that in the former case the etching involves diffusion whereas in the latter case the etching involves collision. Dry etching that employs both physical and chemical processes are known as Deep Reactive Ion Etching (DRIE) or simply RIE. As already mentioned, the DRIE is achieved by bombarding the target (material to be etched) with highly energetic chemically reactive ions. The bombardment causes dislodging of atoms from the material as achieved in sputtering technique. DRIE technique permits the fabrication of high aspect ratio silicon structures. This is why the process is known as deep trench silicon etching process. The sidewall profiles of 90° ± 1° with aspect ratios of up to 40:1 can be achieved. The main benefits of the DRIE are:

- Etch rate of up to 10 μm/min
- Aspect ratio up to 50:1 and sidewall profile nearly perpendicular with tolerance 1 per cent
- Selectivity to positive resist > 75:1
- Selectivity to silicon oxide >150:1
- Etch depth capability 10 to 600 μm
- Feature size 1 to 500 μm

2.8 PROBLEMS WITH BULK MICROMACHINING

Unavoidable problems associated with bulk micromachining process are as follows:

- Extensive real estate (RE) consumption
- Difficulties in etching at convex corners
- Difficult in preparing the mask with high precision
- Etch rate is very sensitive to both agitation and temperature making it difficult to control both lateral and vertical geometries of the structure

2.8.1 RE Consumption

Bulk micromachining is involved with extensive real estate consumption because the process removes a comparable amount of materials from the bulk substrate. Figure 2.20(a) illustrates the fabrication of two membranes created by etching through a (100) wafer from the backside until an etch-stop is reached. A large amount of silicon has been wasted to create this structure (Figs. 2.20(a&b)). The consumption of silicon cannot be stopped but it can be reduced best by practice and careful design as well as employing appropriate etching techniques. The ultimate objective is to achieve the RE gain by reducing the consumption of wafer materials. Two different real estate gain techniques are:

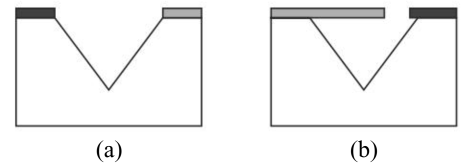


Fig. 2.20 *Wastage of material*

- RE gain by etching from the front
- RE gain by using silicon bonded wafers

RE can be gained by removing fewer amounts of material and the use of thinner bulk instead of thick wafers. This approach is useful for wafers with a thickness more than 200 μm. An alternative solution is to etch the wafer from the front. Anisotropic etchants can remove an amount of materials depending on the orientation of the wafer with respect to the mask, and etch until a pyramidal pit is formed. This produce an angle of 54.7° to the surface of the silicon wafer. The etching stops when the pyramid pit is completed. Hence, there is no further wastage of the wafer.

Another process to gain real estate is using thinner wafers and joining them together using the silicon direct bonding (SDB) technique (described in this chapter). SDB is a process of bonding two silicon wafers by applying pressure and annealing them at a certain temperature. An example is given in Fig. 2.21. A groove can be made using two parts. First, a thin silicon wafer (100) is taken and then it is etched to form a pyramid pit as shown in Fig. 2.21(b). In the second stage it is bonded or fused with another thin wafer to produce another desired structure (Fig. 2.21(c)). Finally, the bonded wafer is etched from the backside of the surface in order to get the final desired shape as shown in Fig. 2.21(d). The RE gain by using silicon fusion-bonded wafers can reduce the consumption of wafers, waste of etchants and time. This process saves RE and to some extent makes the design and etching simple.

2.9 SURFACE MICROMACHINING

Surface micromachining is characterized by the fabrication of micromechanical structures from the thin films of appropriate materials. The films can be composed of materials such as polycrystalline

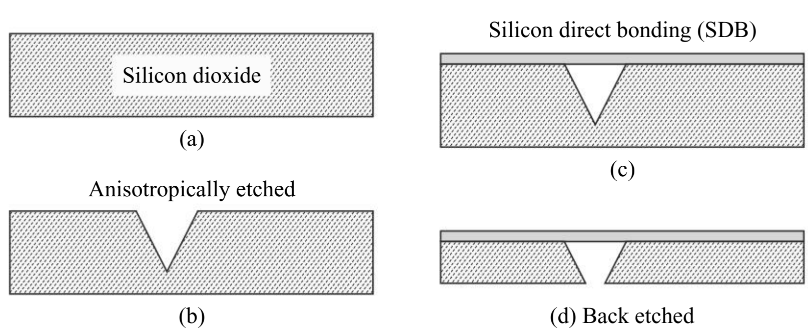


Fig. 2.21 *Real estate gain by silicon diffusion bonding (SDB)*

silicon, silicon nitride, silicon dioxides, etc. and can be sequentially deposited and selectively removed to build 3D microstructures. The essential property of the thin film materials is that they be freed and released from the planar substrate as per required design.

2.9.1 Background

Although surface micromachining fabrication process began in the 1960s, its rapid growth over the past few years is considered significant. Surface micromachining adopts batch fabrication methodology in which large quantities of MEMS devices can be produced simultaneously with low manufacturing cost. The first exemplar surface micromachined device for research study was designed based on resonant device. The configuration of this kind of device is observed to be an x-y plane surface and the dimension was limited to less than 10 micrometer. Because of the nature of the x-y plane and apparent look of such devices the involved micromachining process believed to have been named as surface micromachining. The first application of surface micromachining was reported in 1989. Immediately after that microscale movable parts like joints, gears, springs and many other mechanical and optical components were developed in academic as well as industrial research centers. Recently, many commercial devices are being manufactured using surface micromachining processes.

2.9.2 Design Dependency

The MEMS developers are also interested in material properties such as average strain, strain gradient, Young's modulus, fracture strength, and material damping, since these parameters are functions of the film morphology (study of form and structure), which is in turn dependent upon deposition and/or fabrication processing methods. Figure 2.22 schematically illustrates these interdependencies.

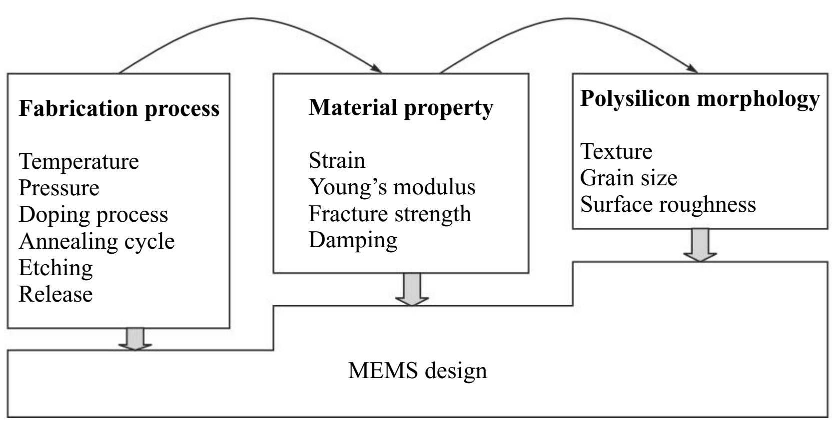


Fig. 2.22 *Microstructural polysilicon mechanical property dependency on film morphology and fabrication process parameters*

Four major issues concern the development of surface micromachining process.

- The understanding and control of the material properties of microstructure films
- Releasing the microstructure from base substrate, i.e. avoiding film stress and sticking.
- The design constraints, i.e. size and layer position.
- The packaging and material constraints, i.e. bonding and safety.

2.9.3 Surface Micromachining Processes

The surface micromachining process sequences are presented in Fig. 2.23. We like to consider silicon as the ground plane substrate because of the reasons mentioned in the first chapter.

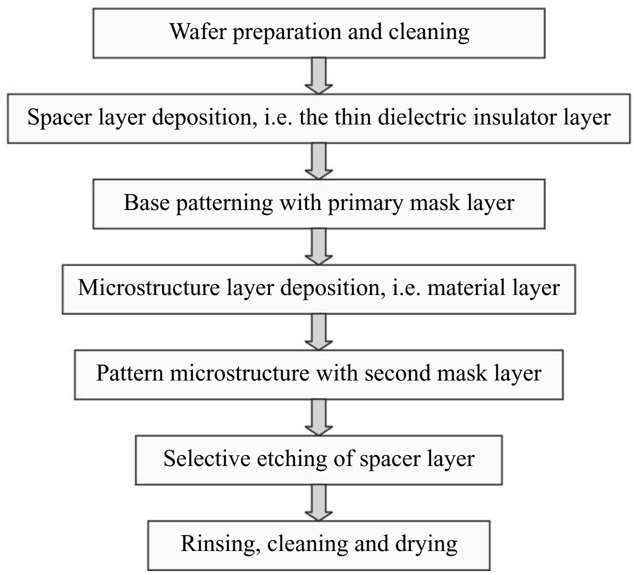


Fig. 2.23 *Surface micromachining process sequence*

As usual the wafer has to be prepared and cleaned. First a spacer layer is deposited on a silicon substrate, which is coated with dielectric materials. The pre-processing is performed prior to the deposition of spacer layer. The spacer layer, called sacrificial layer, is the most important layer in surface micromachining. Sacrificial layer will be removed by etching process at a later stage. One important property of the sacrificial layer is that they must be etched easily and quickly. This layer could be a thin dielectric insulator layer. Phosphosilicon glass (PSG) or SiO_2 is mostly chosen as the material for the sacrificial layer. Addition of phosphorus can enhance the etch rate. Aerogel (described in Section 2.3.2) can effectively be used as sacrificial layer because they are highly porous and have high etch selectivity requiring small etch-time as compared to non-porous materials. The thickness of this layer varies from device to device. On the top of the layer a primary mask layer is coated. This layer is required for specific type of MEMS structure where a window between the sacrificial layer and the material layer (to be deposited next) is needed. This layer is actually needed for the fabrication of complex 3D MEMS structure. The primary mask layer is generally omitted in many structures. Next a thin structural material is deposited either using chemical vapor deposition (CVD) technique or sputtering. The commonly used structural material is polysilicon (poly-Si). Some of the other important structural materials are silicon nitride, silicon oxynitride, polyimide, diamond, SiC, GaAs, tungsten, α-SiH, Ni, W and Al. Silicon nitride and silicon oxide exhibit undesirable high residual stress, which hampers their use as mechanical components. But a mixed silicon oxynitride can produce low or stress-free components. The deposition is achieved using low pressure ranging from 25 to 150 Pa and temperature around 600°C. The normal deposition rate is about 200 Å per minute. The next step is called selective etching (SE). For SE a mask will be needed as mentioned in the section on photolithography (See Section 2.2).

Selective etching is significantly a central process sequence through which foundation for the real free standing of the microstructures is created. Reactive Ion Etching (RIE) process is primarily used for etching the material layer (i.e. microstructure layer). The wafer is then wet etched to remove the sacrificial layer. The structure is finally freed from the spacer layer. Note that etching is to be carried out twice. First, etching is performed on the structural material in order to acquire the desired structure. Second etching is performed on sacrificial layer so that a free structure is obtained. Rinsing and drying process follow etching. A schematic diagram of surface micromachining process is illustrated in Fig. 2.24.

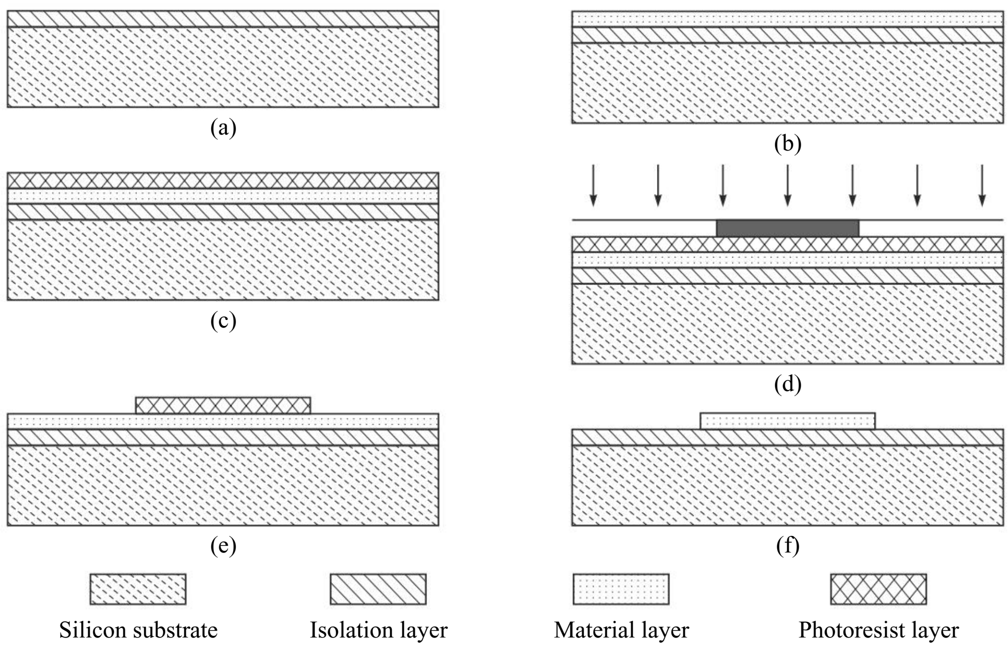


Fig. 2.24 *Surface micromachining (a) Isolation layer (sacrificial or spacer layer), (b) Material layer, (c) Photoresist layer, (d) Photomasking (e) Removal of excess photoresist, (f) Selective etching of spacer layer*

A simple surface micromachined cantilever beams is illustrated in Figs. 2.25(a) and (b). As an example, the polysilicon can be deposited and patterned using RIE technique, followed by wet etching of oxide layer under the beams in order to freeing them from the substrate.

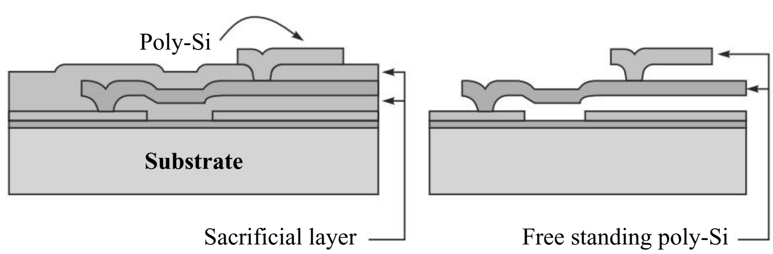


Fig. 2.25 *(a) An example of micromachining cantilever structure, (b) After micromachining (Free standing cantilever)*

It shows that surface micromachining is a powerful technique for producing complicated 3D microstructures. Some of the complex structures which can be designed using surface micromachining process are tweezers, gear trains, micromotors, spatial light modulator (SLM), microphone, gyroscope, and so on.

2.10 BULK VERSUS SURFACE MICROMACHINING

Bulk and surface micromachining techniques differ in many ways (Table 2.9). Bulk micromachining confers the use of Si(100) as starting material, anisotropic electrochemical etchants—each different type of etch stop, and double-sided processing. Bulk micromachining process removes the materials from the substrate in either direction. The process uses larger die and sometimes wastes good percentage of material with high production cost.

Surface micromachining uses dry etching method in patterning and isotropic etchants in release steps. The use of polysilicon in surface micromachining avoids many difficulties associated with bulk micromachining. This increases the freedom for the design and integration of many types of MEMS designs such as sensors, actuators, mechanical switches, and so on. The use of sacrificial layer gives advantage for assembly of tiny mechanical structures. Since surface micromachining normally adds the thin film it is mainly active in x-y plane and therefore difficult to process on z-axis. Table 2.10 describes the difference between the bulk and surface micromachining in a comparison form.

Table 2.10 *Bulk versus surface micromachining*

Bulk micromachining	*Surface micromachining*
Aspect ratio (vertical dimensions): Relatively high (One or more wafer thickness)	Vertical dimensions are limited to the thickness of the deposited layers (~2 μm), leading to compliant suspended structures with the tendency to stick to the support
In most situation the machining involves laminating	Surface micromachined device has its built in support and is more cost effective
Large features with substantial mass, thickness and volume	Relatively small features with low mass, thickness and volume
Not very compatible to traditional IC-based technology	Natural but complicated integration with circuitry; Integration is often required due to the tiny capacitive signals
Piezoresistive or capacitive sensing are easier to implement	Capacitive and resonant sensing mechanisms because of easy formation of air gap and cavity
Sawing, packaging and testing are difficult	No excuse
Substrate are susceptible to fragile near the end of the production	Cleanliness is severely critical near the end of the micromachining process
The technology is relatively matured: Abundance of products and producers	Relatively new technology for which no mature products and producers are available. But building momentum
Utilizes both sides of the substrate; Especially during the process of etching	Several thin film deposition and etching are required to build up 3D microstructures

Source: Adapted from H Jerman, Canada

2.11 WAFER BONDING

Silicon wafer bonding technique is used to join two wafers together. The physics behind the technique is simply the effect of molecular-level bonding. Integrated fabrication is preferable in microdevices, but sometimes due to complex shapes and fabrication difficulties, bonding of two or more parts of the device are preferred. Many bonding methods have already been developed and successfully applied in MEMS devices.

The wafer bonding process can be illustrated as a four-step sequence: surface preparation, fusion, pressurisation and annealing. The first step is important because the quality of the bond depends on the surface conditions. Any surface contamination due to particulate material can damage or create poor bonding. A particle of size 1 μm can have an effect on an area as large as 1 cm in diameter. In order to achieve good bonding effects, typical values of surface roughness and flatness are 5 Å and 5 μm on a 4-inch wafer, respectively.

Bond strength is critical and sometimes difficult to assess in many applications due to microscale dimensions. A number of techniques can be applied to evaluate this. One of the most common techniques is based on a surface energy measurement proposed by the researcher Maszara. A blade of known thickness is inserted between the bonded elements to create a crack. The length of the crack is then measured. The surface energy can be calculated using the imperial formula as given in Eq. 2.6.

$$\lambda = \frac{3}{8}\frac{Et^3y^2}{L^4} \tag{2.6}$$

where, λ is surface energy, E is modulus of elasticity, t is the sample thickness, $2y$ is the blade thickness and L is the length of the crack. The main drawback of this approach is the fourth power on the length of the crack. An error in the measurement is increased in the order of the fourth power.

There are several techniques, which can be used for bonding wafers. Important techniques are:

- Glass-frit bonds
- Anodic and fusion bonds

Typical process conditions used for anodic and glass-frit bonds are presented in Table 2.11.

Table 2.11 *Typical wafer bonding*

Bond type	*Temp. (°C)*	*Pressure (Bar)*	*Voltage (V)*	*Surface roughness (nm)*	*Precise gaps*	*Hermetic seal*	*Vacuum level during bonding (Torr)*
Anodic	300–500	N/A	100 V–1 kV	20	Yes	Yes	10^{-5}
Glass-frit	400–500	1	N/A	N/A	No	Yes	10

2.11.1 Glass-frit Bonds

Glass-frit bonding is a traditional method. Low melting point glasses is used for forming hermetic seals. The process is typically carried out in the temperature range 400–650 °C and contact pressures of ~10^5 Pa. The Motorola automotive air bag accelerometer is the most recent example of silicon wafer bonding (Fig. 2.26). In this microdevice, a silicon top cap is bonded at the wafer level with the triple-

level polysilicon surface micromachined accelerometer using a low-temperature glass-frit bond. This wafer-level silicon cap provides mechanical protection and also prevents damping and shocking.

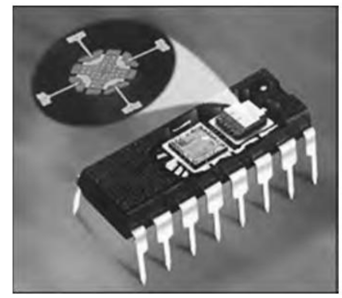


Fig. 2.26 *Wafer bonding application*
Courtesy: Freescale Semiconductor

2.11.2 Anodic Bonding

Anodic bonding is applied to silicon wafers and glass (pyrex borosilicate) with a high content of alkali metals. It has a sodium oxide (Na_2O) concentration of 3.5%. Anodic bonding is also known as electrostatic bonding. In this approach, silicon and glass is arranged as shown in Fig. 2.27. An electrode is connected with the glass. High negative voltage is applied between the two mechanical supports and the combination is heated up to 500°C. The high electric field in the entire area creates a strong electrostatic force, pulling the two surfaces together. The glass contains positive ions (Na^+) attached to the negative electrode area where they are neutralised by the counter charges. The transition creates a space charge at the glass-silicon interface and develops strong electrostatic bonding between them. During bonding, oxygen from the glass is transported to the glass-silicon interface. Because of the movement, SiO_2 is formed which in turn creates a permanent bond. During the process, the electric field is high enough to allow a drift of oxygen to the positive electrode (Si), which then reacts with silicon and creates a Si-O bond. This technique can produce uniform bonds and it has been successfully applied to many microsystem applications including pressure sensors, solar cells and piezoresistive devices. However, this technique is not good for complex actuators.

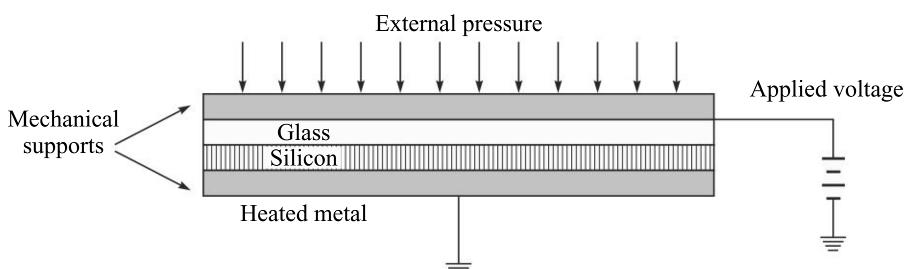


Fig. 2.27 *Anodic bonding arrangement*

It is also possible to bond silicon wafers together by applying gentle pressure. Other bonding methods include using an adhesive layer like wax, epoxy and SiN–SiN bonding for low temperature processes and Au-coated bonding (sometimes called eutectic bonding) using glass or photoresist. Anodic bonding can form very strong joints but they demand that the surfaces be very clean and flat. Wafer bonding techniques can be combined with basic micromachined structures to design complex microdevices like valves and pumps.

2.11.3 Fusion Bonding

Fusion bonding is known as Silicon-Direct-Bonding (SDB). The SDB technique has been used in a wide range of applications in microelectronics and microsystems technology. It is a suitable method for

manufacturing devices including p-n junctions. Two silicon wafers are bonded through the formation of oxides on their surfaces. This technique is commonly used in bulk micromachined MEMS packaging.

Bonding relies on the need for very smooth and flat surfaces to adhere. Maximum bond strength is obtained at temperatures between 700–1100°C. Thermally sensitive devices can be bonded with sufficient strength at temperatures between 200–400°C by using chemical surface activation methods. The first step of low temperature SDB of two wafers is the cleaning and hydrophilisation in an acid mixture (H_2SO_4 and H_2O_2), followed by rinsing and spin-drying. The wafers are then wetted with silicate solutions like sodium silicate (NaSi) or tetraethylorthosilicate (TEOS). The wafers are rinsed and dried for a second time. Then, they are joined by applying an external force or load. Bonding is initiated with a single point contact at the middle, and subsequently a uniform pressure is applied.

Bonding propagates from the centre to both sides and it is influenced by a change in viscosity and pressure of the ambient gas, as well as the surface energy. The mechanical spacers, as shown in Fig. 2.28, are usually provided in order to facilitate mechanical and electrical contacts. Sometimes gases and moisture in between the wafers make the bond weak. For this reason the wafers are heated in order to dehydrate the surface. The wafer molecules diffuse from Si–to–Si bond as the process continues. At higher temperatures, oxygen is also diffused into the crystal lattice.

SDB is a low-cost packaging technology for silicon and silicon compounds such as SiO_2 or Si_3N_4. In comparison to other bonding techniques, SDB has the following advantages:

- Flexible and low cost
- Thermal mismatch between the bond materials is low
- Bond partners can be featured easily
- High strength because of homogeneity
- Can be bonded at high or low temperature depending on the material used

When bonding occurs at temperatures greater than a thousand degree centigrade, the strength approaches that of silicon itself. But this process has shown some drawback in some applications, especially concerning the formation of cavities.

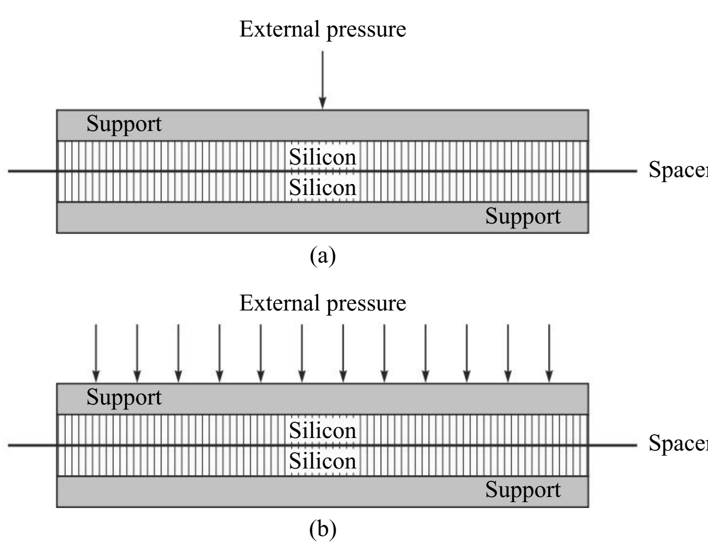


Fig. 2.28 *Fusion wafer bonding, (a) Single point contact, (b) Uniform surface contact*

2.12 LIGA

LIGA is an acronym for the German words[3] for lithography, electroforming, and plastic molding. LIGA is a molding technology that produces precision microsensors, microactuators and microparts for MEMS and MOEMS devices. The three steps mentioned above make it possible to mass-produce microcomponents at a relatively low cost. Normal lithography process is used in first place (Refer to Fig. 2.29). Deep X-ray lithography (DXL) is essentially used DXL allows creation of high aspect ratio microstructures close to vertical dimension of 1 mm and a lateral resolution down to 0.2 μm. DXL is highly penetrating, intense and vertical radiation process. The synchronous X-ray radiation is supplied from a synchrotron. The photomask is prepared by using two types of materials: absorber and the transparent carrier. The thick absorber absorbs the X-rays and the thin transparent carrier allows the X-rays to pass through. The absorbers are usually metals such as W, Ta or Au and the commonly used transparent carriers are Si_3N_4, Ti, Be, Poly-Si, and SiC. CAD tools help in preparing the mask. The photoresist materials are PMMA, polycarbonate, epoxy phenol resins, polyvinylidene fluoride (PVDF), polysulfones and polyether ketones. PMMA (polymethylmethacrylate) is mostly used. The exposing energy then modifies the plastic material (i.e. PMMA) in such a way as to ensure that with a suitable solvent the material is removed, leaving behind the structure of the unirradiated shadowed areas. This phase of the process is called development (Figs 2.29(a), (d) & (e)). After the development the primary structures are formed (Fig. 2.29(e)). Thereafter the electroforming process starts. Electroforming is a

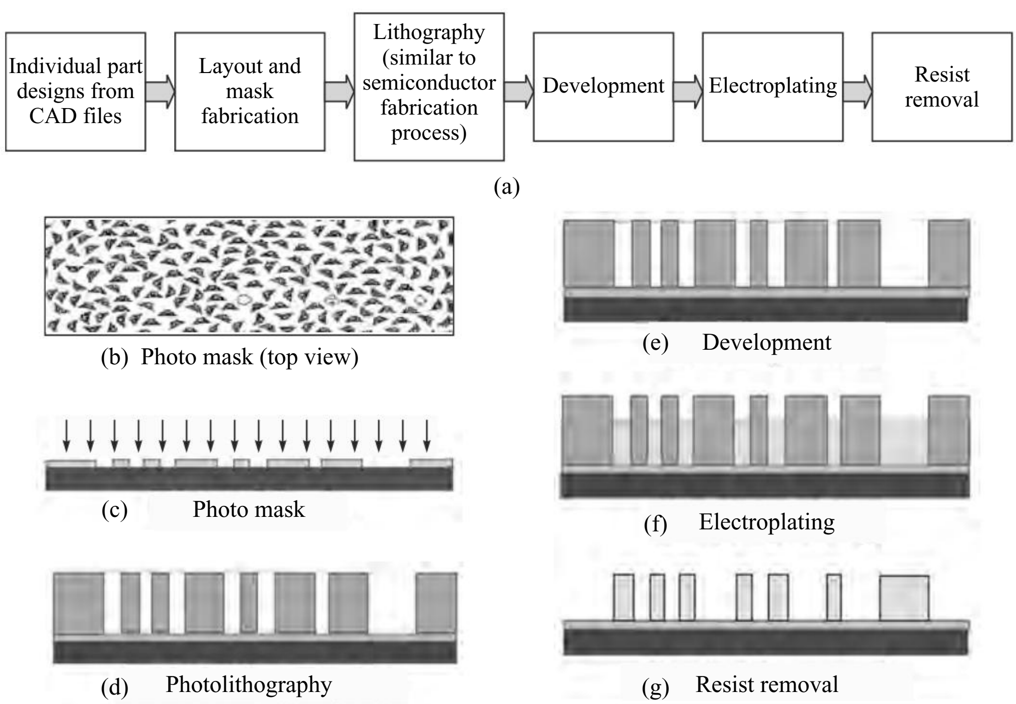


Fig. 2.29 *LIGA process*

3 **LI**thographie—Lithography; **G**alvanoforming—Electroforming; **A**bformug—Molding

process to form metallic microparts on the primary structures. This is achieved in terms of filling the cavities of the PMMA patterns, formed after the deep X-ray lithography by means of metals such as Ni, Ni–Fe, Ni–Co, Au, Cu, Ag. Electroforming is thus a process for creating 3-D metal parts by using a carefully controlled long-duration electroplating process. In this way, the negative pattern of the plastic structures is generated as a secondary structure out of metals. The metal microstructures produced by deep X-ray lithography and electroforming is used as molding tools for the production of replicas of the primary structure in large quantities and at reduced cost. The LIGA technique is thus used to produce microstructures for direct use. Various microstructures thus formed in this way can be assembled to form a desired MEMS device. Usually LIGA process is useful in micromanufacturing MOEMS components and devices, although its applications are found to micromanufacture the following micro-components and systems.

- High frequency antenna arrays
- Cams, Gears, Levers
- Frames, Springs, Sliders, Latches
- Micro probes, Grippers, Cutters, Motors

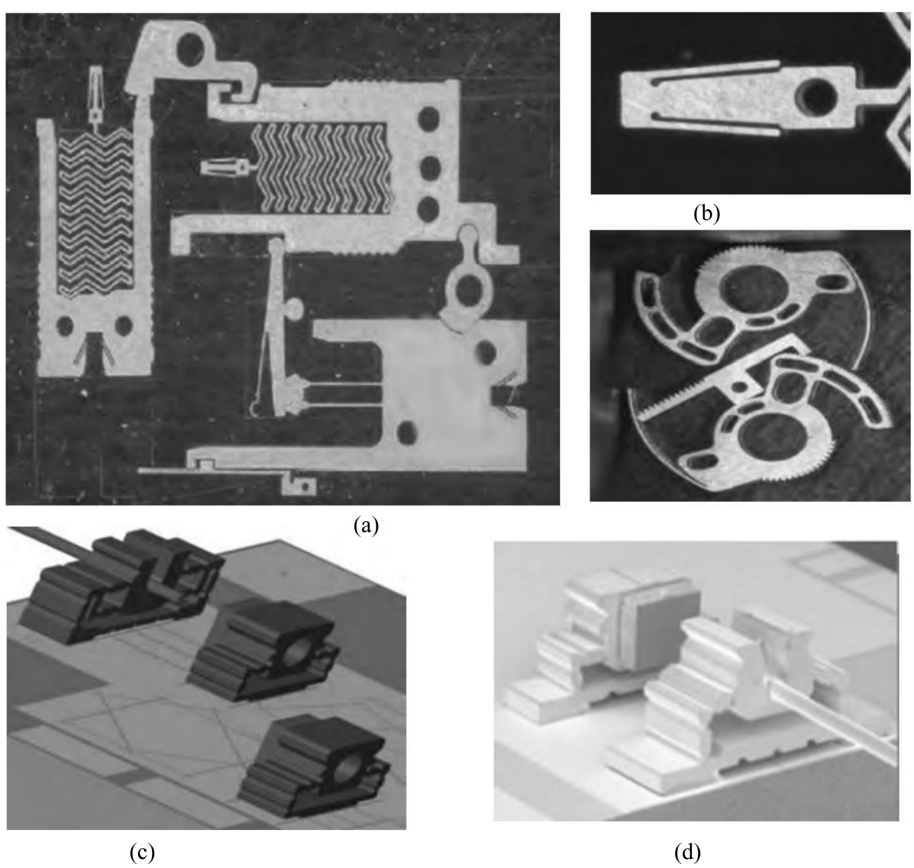


Fig. 2.30 *LIGA fabricated components. (a–c) Various types of passive mechanical systems (PMS), (c–d) Lens and fiber alignment structures*
Courtesy: AXSUN Technologies, Inc.

- Micro-nozzle arrays, Micro-fluidic delivery systems

The LIGA process sequences are summarized below. Figure 2.30 shows a microsystem developed by LIGA process. The precision dimensional range of LIGA processes compared to silicon and precision micromachining is illustrated in Fig. 2.31.

- Preparation of photomask using CAD tools.
- Patterning of a thick photoresist with accurate exposure from synchronous X-ray radiator
- Development of desired structures after the extended exposure
- Electroplating of cavities, space in order to cover up the photoresist up to the desired level
- Removal of photoresist in order to form the complementary or replica metal structure
- Use of metal structure as a mold

Practically many types of shapes can be formed in two dimensions with vertical sidewalls. Typical feature parameters are

Thickness	:	Start from 100 microns to several mm.
Number of microparts per wafer	:	Up to 3000 microparts can be electroplated
Minimum feature sizes	:	15 microns
Aspect ratios (H/W)	:	75/1
Sidewall straightness	:	About 1 micron per mm

AXSUN Technologies is famous the LIGA process. AXSUN Technologies develop, manufactures, and assembles performance enhancing miniaturized optoelectronic and mechanical products for

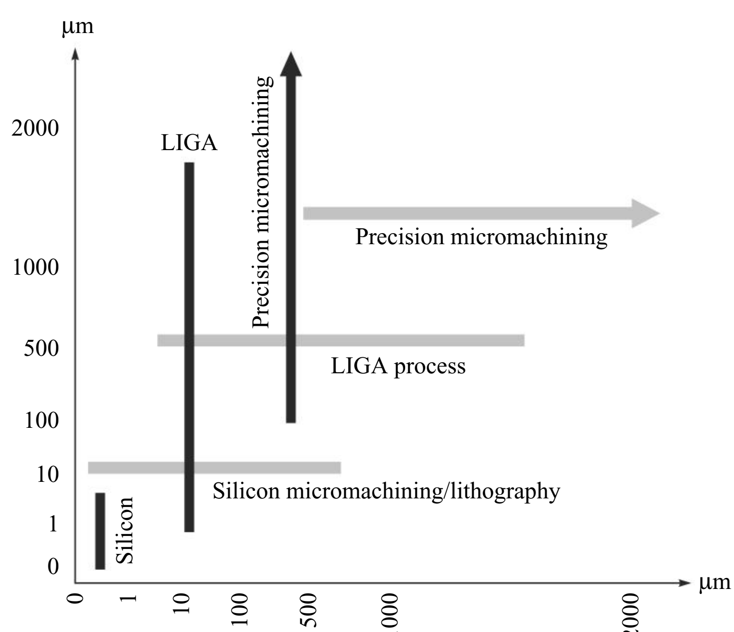


Fig. 2.31 *Comparison of surface and vertical dimension of LIGA with other machining and lithography processes*

communications, defense, life sciences, and industrial applications. The advantages of LIGA process are as follows:

- Rugged, high precision metal parts
- Finished components without micro machining
- Readily assembled to create mechanisms
- Attach by soldering, welding, brazing, or adhesives
- Superior mechanism performance
- High precision metal structures for mounting and aligning micro optical devices
- Enables both active and passive device alignment
- Deformable for precise multi-axis alignments to 0.1 micron
- Precise surfaces for easy, accurate mounting
- A technology for fabricating highly precise micro components from metals and plastics

2.13 SUMMARY

The word micromachining has been derived from machining which is a process of designing mechanical components of various shapes. The design process which deals with the components of the order of micrometer is obviously known as micromachining process. Micromachining not only designs mechanical components but electronic and optical components as well. The design of these microcomponents is known as fabrication. So micromachining is a fabrication process which of two types: bulk micromachining and surface micromachining. In fact, micromachining is an advanced version of the traditional IC (Integrated Circuit) fabrication process. Micromachining mostly considers fabrication of 3D mechanical components along with electronic circuits.

Since micromachining is very compatile to IC fabrication process, this chapter starts with explaining various types of lithography methods, the primary fabrication process used in IC manufacturing. Lithgraphy methods such as UV photolithography, x-ray photolithography, electron-beam lithography, soft-lithography and their comparisons are presented. Various types of photoresist materials used in the lithography process are presented. Regardless of whether process is for IC fabrication or MEMS micromachining, both the design methods require deposition of thin films during manufacturing. In this respect characteristics of thin film and deposition methods such as LPCVD, PECVD, e-beam evaporation and sputtering are presented in order. Etching is an important sub-process in micromachining. Etching is defined as removal of selective materials either from the thin film or from the substrate itself in order to make a desired shape. Etching is achieved in two ways: wet etching and dry etching. Dry etching and wet etching methods with examples of etchants and their comparisons are presented in the appropriate section. With respect to wet etching, we classify it under two categories: anisotropic and isotropic etching. Etching is controlled by using etch stop. Introduction to etch stop is given. In particular, doping based etch-stop mechanism is suggested because of advantages pointed out. The doping can be achieved by diffusion and ion-implantation method. The last part of the chapter draws comparisons between bulk micromachining and surface micromachining process. Important surface micromachining process parameters are listed. The chapter ends with introducing wafer bonding method, another process involved in MEMS micromachining.

Points to Remember

- Machining is defined as the process of removing material from a work piece in the form of chips in order to obtain the exact shape and size of the work piece that is required.
- Micromachining is considered a process as well as a technology that is utilized to structure wafer materials or thin films in order to fabricate miniature devices such as microsensors, microactuators and passive components (electronic circuits, post, gear, hinges, flexure, etc.) for microsystems functioning as electromechanical, optoelectronic and optomechanical systems.
- If structuring is performed on wafer materials the process is called bulk micromachining process and if it is performed on the thin film the process is called surface micromachining process.
- Lithography is a Latin word which means stone-writing. Stone-writing is a process of printing that utilizes flat inked surfaces to create the printed images. Photolithography is an optical means of transferring patterns into the substrate.
- The foremost phase of the photolithography is the preparation of a *photomask*.
- The UV light sensitive polymer material is called photoresist. There exist two types of photoresist; positive and negative photoresist.
- The photoresist is coated on the surface by a process called *spin-coating*.
- Etching is simply considered as removal of selective materials from the wafer.
- The materials from which the microcomponents are built are called structural materials.
- Sacrificial layer is a layer, which is deposited on a specific region and later on is removed so that microstructures or features can be elevated.
- SU-8 is an epoxy-resin negative photoresist widely used for fabricating MEMS structures. The advantage of using SU-8 photoresist is that high a*spect ratio* of approximately 20:1 3D microstructures can be built.
- Aerogel is virtually weightless solid measuring 0.00011 lbs. per cubic inch. Low density makes it useful as a lightweight structural material, and its super-high internal surface area makes it a super-insulating solid material.
- Other lithography methods are X-ray, electron beam soft-lithography.
- Photolithography is a hard-lithography method. The last part of this section deals with soft-lithography method.
- X-ray lithography uses collimated X-rays as the exposing energy.
- The interaction of matter with light and X-ray differ. As a consequence, compatible mask materials are chosen for X-ray lithography.
- Kapton® polyimide film is primarily used for insulation purpose.
- E-beam lithography uses a focused electron beam to write ultrafine patterns. The beam can take the shape of square or rectangle. E-beam method can produce structure with better resolution compared to other lithography method.
- Soft lithography is another techniques for the fabrication of micro- and nanostructures. It is a relatively recently developed technique that can also facilitate high-resolution patterning and molding of soft polymers materials effectively.
- PDMS block with patterned relief structures on its surface is the key to soft lithography and can be used in diverse processes requiring patterning.
- The elastomeric mold is prepared by cast molding.
- The thin films can form microstructure, conducting, and insulating or even semiconductor layers.
- Film deposition can be achieved using spin coating, thermal oxidation and chemical vapor deposition (CVD), e-beam evaporation and sputtering.
- Thermal oxidation is classified either as *dry oxidation* or *wet oxidation* depending upon which oxidant species is used.

- LPCVD (Low Pressure Chemical Vapor Deposition) is a technique in which one or more gaseous form of materials are used to form either a film (could be an insulating or conducting) on another surface such as a wafer under low pressure and relatively high temperature conditions. LPCVD creates thin films of material on a substrate via the use of chemical reactions.
- Much like LPCVD, PECVD is used to form a solid film on the surface of a wafer by the use of one or more gaseous reactors. Exclusively, in this case the operating temperature is much lower than the LPCVD process.
- The electron beam evaporators are used to deposit thin films onto substrates and are faster compared to other depositing techniques. Evaporation is achieved by locally heating a source material above its vapor pressure with an electron beam.
- Sputtering is a process by which atoms are extricated from the surface of a target source material as a result of collision with high-energy atom or ion particles generated from plasma setup. The phenomenon can be exploited for developing thin film onto a substrate.
- Sputtering has been one of the most widely used techniques for depositing various metallic films on wafers, including aluminium, aluminium alloys, platinum, gold, TiW, tungsten, etc.
- There are two ways of accomplishing sputtering mechanism; diode sputtering and magnetron sputtering.
- Magnetron sputtering technology overcomes the wasted electron problem.
- Sputtering is also possible on inorganic materials.
- The process of adding impure material to a material of interest is called *doping* and the impure materials involved in this process are called *dopants*.
- Dopants are of two types, *donors* and *acceptors*.
- Doping is also an essential process in MEMS designs. Doping makes it possible to modify the properties of the material.
- There are two types of doping process, *diffusion* based doping and *ion implantation* based doping.
- When a calculated amount of specific dopants are carefully diffused into the desired material in a controlled manner to achieve the required properties, then it is called diffusion based doping.
- Diffusion occurs when evaporated dopants are allowed to pass on the surfaces of the sample material (could be a wafer) that are placed in a chamber containing inert gas.
- Accelerating the dopant ions in a strong electric field and directing them towards the surface of the material of interest achieve *ion implantation* doping.
- Etching is a process, which makes it possible to selectively remove the deposited films or parts of the substrate in order to prepare desired patterns, shapes, features, or structures.
- Two classes of etching process are common.
- They are *wet etching* and *dry etching*. Wet etching removes the material selectively through chemical reaction.
- Dry etching sputter the material using reactive ions or a vapor etchant.
- The etchants are usually a mixture of acidic solutions.
- The dissolution of material due to chemical reaction may not be uniform in all directions. This characteristic of etching is called directionality.
- Wet etching is classified into two categories depending on whether the process etches the material anisotropically or isotropically.
- In anisotropic materials, the etch rates are not same in all directions. Anisotropic etching is considerably a highly directional etching process with different etch rates in different directions.
- In isotropic etching, materials are removed uniformly from all directions and it is independent of the plane of orientation of the crystal lattice.
- HNA etchant is an autocatalytic solution.
- There are many techniques by which the etch rate can be controlled and monitored.
- Dry etching does not utilize any liquid chemicals or etchants to remove materials.

- Dry chemical etching is also known as plasma etching, while physical sputtering is known as ion beam milling. The dry etching that employs both physical and chemical processes are known as reactive ion etching (RIE).
- Bulk micromachining process removes the materials from the substrate in either direction.
- Bulk micromachining is involved with extensive real estate consumption because the process removes a comparable amount of materials from the bulk substrate.
- SDB is a process in which two silicon wafers are bonded by applying pressure and annealing at a certain temperature.
- Surface micromachining is characterized by the fabrication of micromechanical structures from the thin films of appropriate materials.

Exercises

1. Provide a suitable definition of machining. How are traditional machining processes performed? Comment on how micromachining process differs from the traditional machining.
2. What common methods, principles and process sequences do the IC and MEMS fabrication have?
3. Briefly differentiate the process of bulk micromachining and surface micromachining.
4. What does the photolithography process do?
5. What do you mean by a wafer? What kind of material is commonly used for wafer? Is there any difference between wafer and substrate?
6. What is the role of a photomask in a typical photolithography process? How are photomasks prepared?
7. How many types of photomasks do you know of?
8. What is exposing energy in a lithography process?
9. What is the role of spin-coating and how is it performed? With suitable assumption, write down the relationship between the desired thickness of a thin film and the time of spinning.
10. With suitable schematic diagram and using two types of photoresists, describe in detail the process of photolithography.
11. What are the purposes of having wafer, structural material and sacrificial material in a typical MEMS fabrication?
12. Write notes on the following.
    (a) SU-8
    (b) Aerogel
13. Discuss the following lithography methods.
    (a) X-ray lithography
    (b) E-beam lithography
    (c) Soft lithography
14. Besides spin-coating, what other methods are adopted in order to deposit thin films for MEMS applications?
15. Discuss the following thin film deposition methods.
    (a) Thermal oxidation (b) LPCVD

(c) PECVD
(d) E-beam evaporation
(e) Sputtering

16. List the functional units of a typical PECVD (Plasma Enhanced Chemical Vapor Deposition) system and explain the functions of each unit.
17. What do you mean by wasted electrons in a typical sputtering process? How can wasted electron problems be solved by employing magnetron sputtering? Discuss the relative merits and demerits of diode and magnetron sputtering.
18. What are the advantages of sputtering based thin film deposition?
19. What do you mean by impurity doping? Give an example with respect to semiconductor device. How can doping modify the physical parameters of the material? Show some examples with experimental data.
20. What do you mean by diffusion? Why is diffusion needed? What methods are adopted in order to achieve the process of diffusion? List the process sequences of ion implantation based diffusion method.
21. Define etching. Etching is an essential sub-process in micromachining—Justify.
22. Write notes on the following.
    (a) Wet etching
    (b) Dry etching
23. Distinguish between the anisotropic etching and isotropic etching in the wet etching process. What do you mean by directionality?
24. What do you mean by etch stop? How can etching be controlled? Give some examples.
25. List out the disadvantages of wet etching process.
26. How is dry etching achieved?
27. Write down the reaction steps that occur in dry plasma etching.
28. Write notes on surface micromachining.
29. Distinguish between bulk and surface micromachining processes. What do you mean by RE consumption?
30. Write a short note on wafer bonding.

Chapter 3

System Modeling and Properties of Material

Objectives

The objective of this chapter is to study the following.

- The need of system modeling
- To list the important steps for analysis and design of engineering systems
- To understand the meaning of the term *system* and discuss the types of systems such as SISO, SIMO, MISO and MIMO systems
- To discuss the modeling of a mechanical structure using spring elements
- To discuss the basic modeling elements of the basic systems for mechanical, electrical, fluid, thermal system
- Physical, mechanical, electrical and thermal properties of materials
- Definitions of various useful parameters which describe the properties of materials

3.1 INTRODUCTION

System modeling is about solving practical problems using mathematical models, called model equations, which are formulated looking at the real physical system. An example with respect to a parallel plate capacitor (PPC) type electrostatic MEMS actuator was given in the Section 1.18 in Chapter 1. The important phases of modeling always start with the identification and description of the physical systems and then formulation of a mathematical model. The systems are identified and described utilizing '*basic modeling elements*'. In other words, the basic modeling elements depict a system. The basic modeling elements facilitate the framing of balancing model equation keeping in view the law conservation of energy. Once the physical system is fully described, it becomes easier to study, analyze, evaluate and predict its behavior and performance so that design requirements can be met.

Thus modeling allows us to learn a great deal about the system in terms of understanding and predicting its behavior. As indicated, the law of conservation of energy takes part in the modeling process. All systems including MEMS are understood as the composition of basic modeling elements. Modeling of MEMS systems plays an important role in analysis and optimization. Analysis and optimization is taken to mean the manipulation of system variables and coefficients (together, called

parameters) that are present in the model equation. The results of the analysis help to show how well the system met the design parameters and how best to alter and modify the design as needed. Note that system identification and analysis are the two important phases of system design. The entire design procedure is summarized below:

- Identification and description of the physical system in terms of basic modeling elements.
- Formulation of a mathematical model
- Analysis of the model: Iterative evaluation in terms of achieving effectiveness of the system
- Verification: That is whether all the functional design specifications are met
- Refinement: Propose modifications for any improvement
- Practical implementation: Real design

More description on system simulation will be presented in Chapter 14. However, this chapter deals with fundamentals of basic modeling elements with regard to identification of the physical systems and formulation of mathematical model.

The last part of the chapter introduces material properties.

3.2 THE NEED FOR MODELING: AN EXAMPLE WITH MACRO SYSTEMS

Consider a macro system such as an automated motorized spindle machine. Spindle systems are primarily employed for machining, micromachining, milling, drilling and data reading/recording applications in aerospace, heavy industry, consumer industry, and in many manufacturing sectors. The spindle system is a mechatronic system because they are integrated with electric motor, electronic interfacings and control systems. The electrical motor is an electromechanical device, which contains mechanical and electrical parts. The mechanical parts are the shaft of the motor, the stator body, the armature core, etc. The electrical parts are armature coils and field coils.

There are many issues to be encountered while designing the mechanical and electrical parts. While designing the shaft of the spindle, for instance, the specification parameters such as stiffness, damping, inertia, etc. are to be considered in order to deal with the risk factors such as vibration, loading effect and reliability (possibility of failures such as crack and bending). The geometry of the shaft is considered as one of the design parameters since small deviation could cause a large inertial and vibration effect. Similarly, optimal design is necessary as far as electrical parts are concerned.

What is important to note is that to design a *performance driven* spindle system the optimization fundamental always relates to modeling. Modeling is simply a mathematical description of the physical system. It governs the fundamental physical laws associated with the system. Through modeling a physical systems can be described by mathematical equations called model equations. Since the system is now viewed as the model equations, we are free to manipulate the parameters embedded in the model equation that satisfy the physical laws so that desirable design specifications can be achieved. In effect, modeling helps us to provide sufficient and necessary information to deal with the issues as far as optimization of the system design is concerned. The parameters in our typical example of spindle system could be input voltage, output speed, length of shaft, diameter of the core, the thickness of air gap between the stator and rotor, and so on. The interactive evaluation of parameters in order to determine the effectiveness of the system is called analysis. The results of the analysis help to understand how well the system met the design parameters and how effectively to alter them to get best

design requirement. Then another model is developed for analysis to improve performance. The process continues to iterate the analysis until we have design parameters that satisfy the desired specifications.

3.3 SYSTEM TYPES

Within the engineering arena the word *system* is frequently used. In some generalized sense a *system* is defined as a group of properly arranged *elements* which act together to provide the *desired* output with respect to *available* inputs. Evidently, a system takes some inputs and provides some outputs while satisfying the law of conservation of energy.

One way of classifying the systems is based on their input and output strength. For instance, when the number of input to the system is one and the number of output is also one then such system is called Single Input and Single Output or simply SISO system. Systems with other combinations such as Single Input Multiple Output (SIMO), Multiple Input Single Output (MISO) and Multiple Input Multiple Output (MIMO) also exist (Fig. 3.1). The inputs and outputs can be in different form (For example, the input to an electrostatic actuator is voltage whereas the output is displacement), however as far as energy is concerned, the input energy is equal to the output energy provided that the system is an *ideal* one. In an ideal system, no losses are encountered.

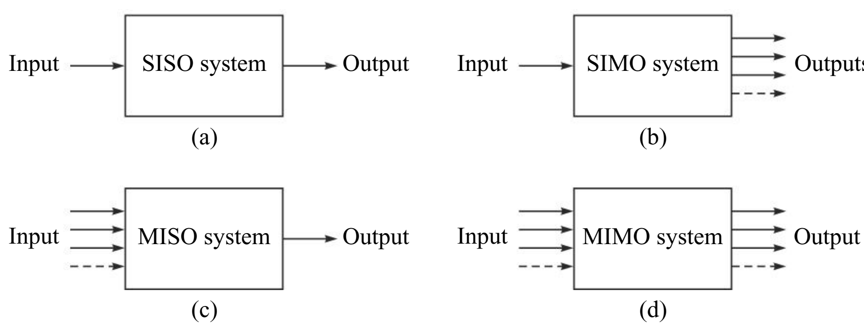


Fig. 3.1 *One way of classifying system*

All systems are categorized under four main basic or elemental systems, namely:

- Mechanical system
- Electrical system
- Fluid system
- Thermal system

The inputs are the sources of energy, which are *available*. The noise is the unwanted energy that enters into the system at any time. The noise energy is considered as the *unwanted available input energy*. So the *total input energy* is the *available input energy* plus unwanted available input energy. The output energy is the difference between the total input energy and the loss. Figure 3.2(a) illustrates a schematic diagram of a system, taking into account all energy components. A system comprising all the inputs, outputs, noise and losses is rather called a ‘dynamic process’. The block diagram of the system and process look similar as can be observed from Fig. 3.2(b). The mathematical description of a dynamic process is expressed as follows.

$$\boldsymbol{Y} = f(\boldsymbol{U}, \boldsymbol{N}, \boldsymbol{X}, \boldsymbol{P})$$

Where, $\boldsymbol{U}(t)$ and $\boldsymbol{Y}(t)$ are measurable input and output signals, $\boldsymbol{N}(t)$ is disturbance signal (noise), $\boldsymbol{P}(t)$ is slowly varying process parameters and $\boldsymbol{X}(t)$ time dependent process state variables and are non-

measurable parameters. The faults make a change in $\boldsymbol{P}(t)$ and in $\boldsymbol{X}(t)$ to produce $\boldsymbol{P}(t) + d\,\boldsymbol{P}(t)$ and $\boldsymbol{X}(t) + d\,\boldsymbol{X}(t)$, respectively.

The inputs of a system are obtained from what is known as the 'source'. The sources are the energy supplying entities. They supply energies in various means. Various means of supplying energy from the energy sources of the above-mentioned systems are described in Table 3.1.

Table 3.1 *System and their energy sources*

Systems	*Means of supplying energy*
Mechanical	Force, Velocity
Electrical	Current, Voltage
Fluid	Fluid flow, Pressure
Thermal	Heat flow, Temperature

3.4 BASIC MODELING ELEMENTS IN MECHANICAL SYSTEM

Mechanical systems are the important basic building blocks of the MEMS systems. From the modeling schematic and from the point of view of law of conservation of energy, distinguishably three *basic modeling elements* are assigned for the mechanical systems. They are,

- Spring
- Damper
- Mass/Inertia

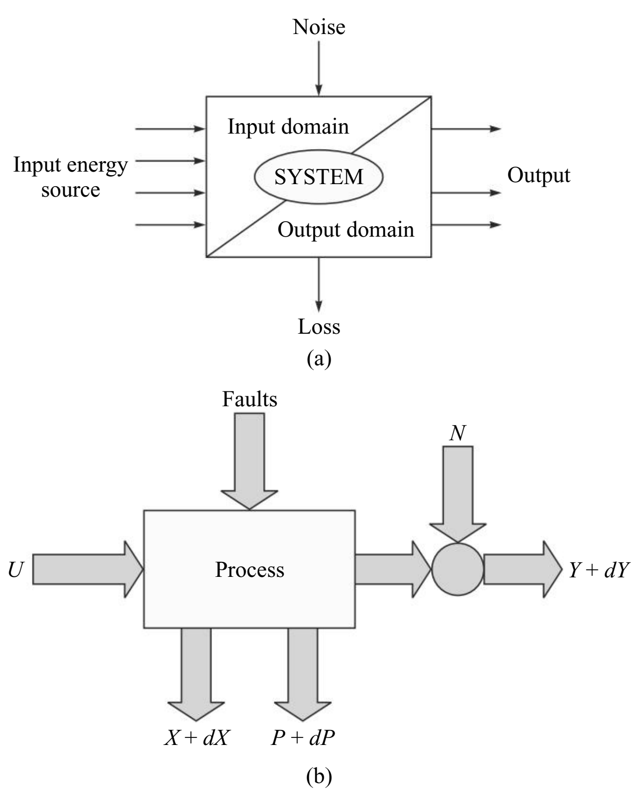


Fig. 3.2 *(a) A schematic diagram of the system, (b) A process influenced by faults*

The fundamental comes from the fact that any kind of mechanical system has spring property, i.e. when force is applied it elongates and the energy is stored within the system. It has also damping property, i.e. when force is applied some portion of the force is lost. Finally it has mass or inertia property that determines how much acceleration it would produce when exerted by the force.

Besides the mechanical system the other elemental systems such as electrical, fluid or thermal system, can also be depicted with similar types of pure elements which are collectively responsible for governing the principle of conservation of energy in their static as well as dynamic states. Those respective elements (described in detail in the sequel) are fundamental because each element characterizes a specific form of energy, which are encountered within the systems. For instance, the spring element stores potential energy, the damper represents the dissipating energy or loss in the system and the inertia element stores kinetic energy. Sometimes, one of these storing phenomena predominates over the others. Then only one of these elements might be sufficient for modeling such a system. In other examples, two or all three elements representing the energy entity might be taken into consideration and in some systems several elements are combined together in order to form the model of the system.

3.4.1 Spring

Spring element stores potential energy. The law that a spring is being described with regard to the storage of potential energy is based on Hook's law. If a force, F (Newton; $kg \times m/s^2$), is applied on the spring as shown in Fig. 3.3, then the spring elongates and the amount of elongation is proportional to the applied force. The mathematical model equation that describes the basic modeling element is called *elemental equation*.

$$F = kx \quad \Rightarrow \quad x = \frac{F}{k} \tag{3.1}$$

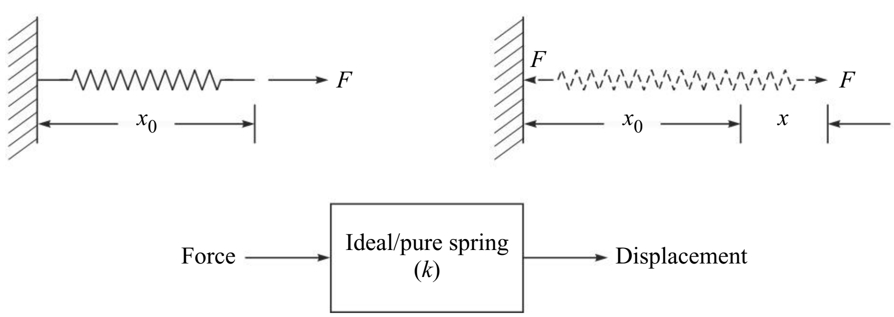


Fig. 3.3 *The spring element (Potential energy storage element)*

Equation 3.1 is called the elemental equation of the spring element. Where, F is the applied force or more appropriately the force exerted by the spring element, x is the change in length or displacement caused by force and k is a proportionality factor called *spring constant* or *stiffness*. The equation implies that if the input to a spring element is force then the output is displacement. The spring element is sometimes referred to as stiffness element. Stiffness is a kind of transfer function. The value of output depends on the transfer function. The relationship between input force and output displacement for spring depends on the geometry and property of the material. The unit of stiffness is Newton/meter. The reciprocal of the stiffness is called mechanical capacitance or *compliance*.

The potential energy (Joules) that the spring element stores is given in the following equation.

$$E = \int_0^x F dx = \frac{1}{2} kx^2 \tag{3.2}$$

The energy can also be expressed in terms of force, as shown in the following equation.

$$E = \frac{1}{2k} F^2 \tag{3.3}$$

A mechanical structure shown in Fig. 3.4 has been modeled using spring element.

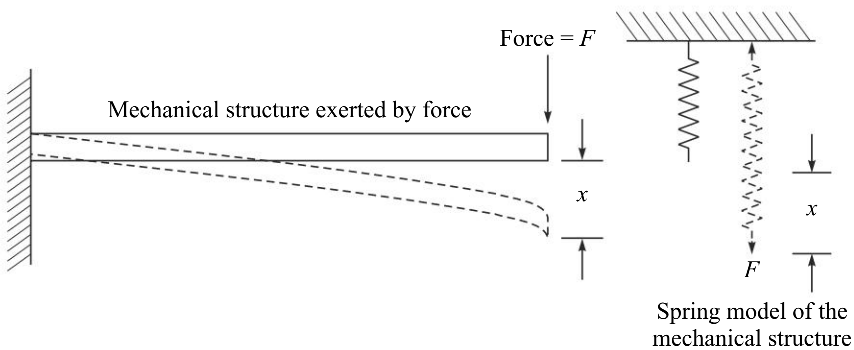


Fig. 3.4 *A typical simple example of modeling a mechanical structure using spring element*

The spring element we discussed is a translational spring since the input is a force not torque. Equally, we have *rotational spring* in which the input is torque. Accordingly, the *rotational spring constant* (or rotational stiffness) should be defined. In this situation, if a torque, T (N-m) is applied, the rotational spring makes angular displacement, θ. That is,

$$T = k\theta \quad \Rightarrow \quad \theta = \frac{T}{k_a} \tag{3.4}$$

Equation 3.4 is the elemental equation of the rotational spring element. Where, k_a is the rotational stiffness (Newton-meter/radian). The energy stored in a rotational spring is,

$$E = \frac{1}{2} k_a \theta^2 = \frac{1}{2k_a} T^2 \tag{3.5}$$

Example 3.1 If the force to stretch a spring is given as $F = (100\ \text{N/m})x$, then what is the potential energy of the spring if it is stretched 5 microns from rest?

Solution: Here, $k = 100$ mN/m $= 100 \times 10^{-3}$ N/m,

and $x = 15$ micrometers $= 15 \times 10^{-6}$ meter.

Therefore, potential energy stored in the spring is,

$$(1/2)kx^2 = (1/2)\,(100 \times 10^{-3})\,(15 \times 10^{-6})^2 = 11.25 \times 10^{-12} \text{ Joule}$$

Example 3.2 It takes a force of 20 millinewtons to hold a spring stretched to a distance of 40 micrometer. What is the elastic potential energy of the spring at this position?

Solution: We know that

$F = 20$ millinewton (20×10^{-3} N) when $x = 40$ micrometer $= 40 \times 10^{-6}$ m.

Since $F = kx$, then $k = F/x = (20 \times 10^{-3})/(40 \times 10^{-6}) = 0.5 \times 10^{3}$ N/m.

Now, elastic potential energy $= (1/2)kx^2 = (1/2)\,(0.5 \times 10^3)\,(40 \times 10^{-6})^2 = 0.4 \times 10^{-6}$ Joule.

3.4.2 Damper Element

Within the system loss is inevitable. The input energy minus the output energy is considered as loss energy, or simply loss. In order to follow up the principle of conservation of energy a loss element should exist in order to describe the system completely, i.e. the energy must be balanced. Indeed, the loss is reflected through the *damper element*. Damper element does not store any energy. It consumes energy, which cannot be recovered. Other name of the damper is *dashpot* that symbolizes resistance, more appropriately *mechanical resistance*. Two types of dampers such as *translational dampers* and *rotational dampers* are defined. Translational dampers and rotational dampers are essentially related to translational movement and rotational movement, respectively. The symbols of the two types of dampers are shown in Fig. 3.5.

The loss is related to the velocity. In translational damper element the relationship between the input force and velocity is linear as expressed in Eq. 3.6. This equation is the elemental equation for the damper.

$$F = b_t v \tag{3.6}$$

where b_t is called *damping constant* (N-s/m). The damper dissipate energy in the form of heat. The power loss can be expressed from the elemental equation, which is given by,

$$\wp = Fv = b_t v^2 \tag{3.7}$$

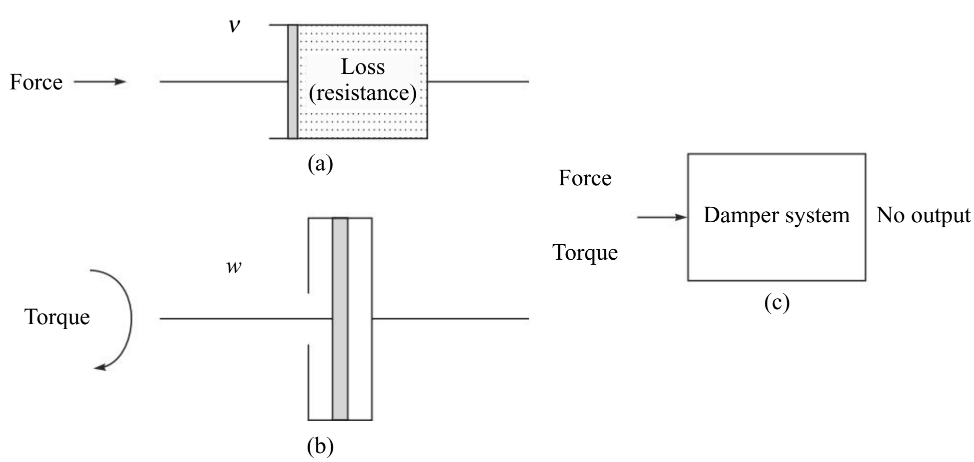


Fig. 3.5 *(a) Translational damper/dashpot, (b) Rotational damper/dashpot, (c) Schematic diagram of a damper system*

The elemental equation of the rotational damper element is,

$$T = b_r w \tag{3.8}$$

where, T is the torque applied, b_r is rotational damping constant in *N-m/s* and w is the angular velocity. The power dissipated by the rotational damper element is then,

$$\wp = Tw = b_r w^2 \tag{3.9}$$

3.4.3 Mass/Inertia Element

Mass element and *inertia element* correspond to translational movement and rotational movement, respectively (Refer Fig. 3.6). Both mass and inertia elements govern the law in terms of storing kinetic energy.

If the body is constrained to rectilinear translation, its body can be considered as a *mass point*. The principle of compatibility is usually applied. Under this situation, the effect of external forces applied to the body is opposed by the body mass and is proportional. Newton's laws of motion are applied while formulating the elemental equation. Newton's second law of motion states the following.

$$F = ma = m \frac{dv}{dt} \tag{3.10}$$

where F (Newton) is the input force acting on the mass point, m (kg) is the mass, and $a = dv/dt$ is acceleration (m/s^2), v is velocity (m/s) with respect to a non-accelerating reference frame. The kinetic energy, E_k stored in the element is given by,

$$E_k = \frac{1}{2} mv^2 \tag{3.11}$$

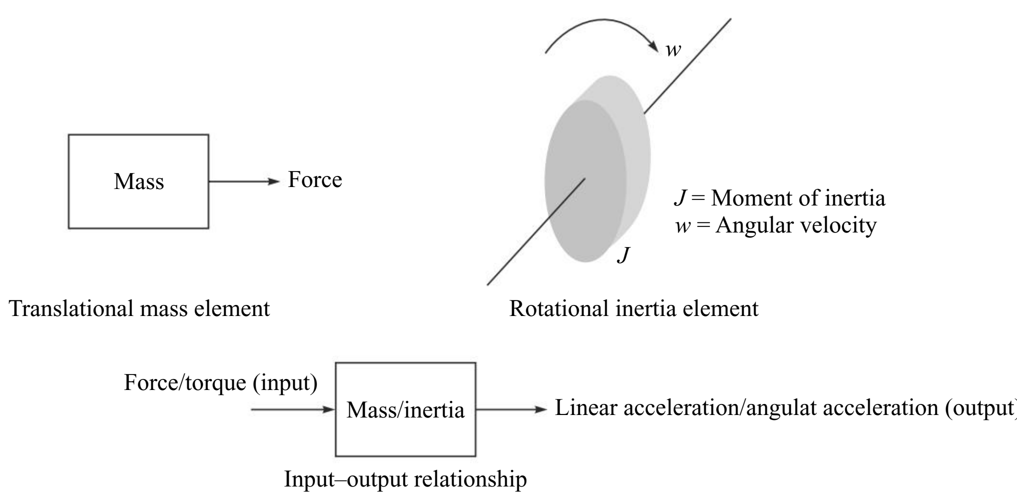


Fig. 3.6 *Mass/Inertia element*

Similarly, with regard to inertia element, the source input torque T is involved in formulating the elemental equation.

$$T = Jw = J \frac{d\theta}{dt} \tag{3.12}$$

where, J is called moment of inertia, $w = d\theta/dt$ is the angular velocity, θ is angular displacement. Kinetic energy stored in the element when it is rotating with an angular velocity w is,

$$E_k = \frac{1}{2} Jw^2 \tag{3.13}$$

System model based on the elements described above (spring, damper and mass/inertia) is referred to as *lumped-parameter* model due to the reason that each element is basic and independent of other. Lumped parameter modeling concept signifies that the system is made up of discrete but pure

components. For instance, when the spring is stretched every bit of ingredient in the spring gets deformed and collectively they (every bit) give the displacement 'x'. The spring is seen as a single unit, a 'lump'. It is not cared how much each part of the spring is stretched due to applied force and how all those every bits add up to the total amount 'x'.

3.5 BASIC MODELING ELEMENTS IN ELECTRICAL SYSTEMS

This section is intended to present the modeling elements of electrical systems. Supporting the fundamental law, electrical systems are also modeled using three basic elements. As before, the appearance of three basic elements has come from the fact that we deal with three types of energy, such as potential energy, loss and the kinetic energy.

This time, however, the elements are

- Capacitor (Spring)
- Resistor (Damper)
- Inductor (Mass/inertia)

(a) (b) (c)

Fig. 3.7 *Basic modeling elements of electrical systems; (a) A capacitor (b) A Resistor (c) An inductor*

The symbols of these elements are shown in Fig. 3.7.

The stored electrical energy, the loss dissipated and the stored magnetic energy in the respective basic electrical elements are expressed in Eq. 3.14 through 3.16, respectively.

$$E_{\text{capacitor}} = \int_0^q v\,dq = \frac{1}{C}\int_0^q q\,dq = \frac{1}{2C}q^2 = \frac{1}{2}Cv^2 \tag{3.14}$$

$$\wp_{\text{resistor}} = \frac{v^2}{R} = i^2 R \tag{3.15}$$

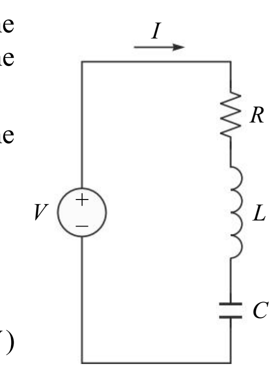
$$E_{\text{inductor}} = \int_0^{\Phi} i\,d\Phi = \int_0^{\Phi} \Phi\, d\Phi = \frac{1}{2L}\Phi^2 = \frac{1}{2}Li^2 \tag{3.16}$$

where,

$E_{\text{capacitor}}$ = Electrical energy; $\wp_{\text{resistor}}$ = Power loss; E_{inductor} = Magnetic energy; i = Current in Ampere; Φ = Flux linkage in Weber; L = Inductance in Henry; v = Voltage across the terminal in Volt; R = Resistance in Ohm; C = Capacitance in Farad; and q = Electric charge in Coulomb.

The values of C, R and I are associated with geometry of the system (Refer Table 3.2).

Example 3.3 Consider the following RLC electrical system, in which all the basic modeling elements are connected in series. Derive the dynamics of the system if the applied voltage is a time-varying voltage.



Solution: In the above circuit, the three components are all in series with the voltage source. Let us define the following.

v—the voltage of the power source (measured in volts V)
i—the current in the circuit (measured in amperes A)
R—the resistance of the resistor (measured in ohms = V/A);
L—the inductance of the inductor (measured in henries = H = V·s/A)
C—the capacitance of the capacitor (measured in farads = F = C/V = A·s/V)

Table 3.2 *The values of C, R and I depend on the geometry*

Basic element	*Geometrical relation*	*Expansion*
Capacitor	$C = \dfrac{\varepsilon_r \varepsilon_0 A}{d}$	where A is area, d is distance, ε_r is relative permittivity of the dielectric material between the plates and ε_0 is the permittivity of the free space.
Resistor	$R = \rho \dfrac{l}{A}$ $\left(\sigma = \dfrac{1}{\rho}\right)$	R = Resistance, ρ is defined as resistivity, in Ohm-meter ($\Omega - m$), which is simply the resistance per unit length times cross-sectional area. A is the cross-sectional area of the material through which the current flows, in meter squared (m^2) and l is the length of the material, in meter (m) and σ is the conductivity of the material.
Coil Inductor	$L_{\text{coil}} = \dfrac{\mu N^2 A}{l}$	where, μ is the permeability of the material through which the magnetic flux passes, N is the number of turns present in the coil, l is the length and A is the cross-sectional area of the coil-turn.

Now, the solution for the current (I) can be obtained by using Kirchhoff's voltage law, i.e. the voltage drop across each element is equal to the supplied voltage. This can be written as,

$$v_R + v_L + v_C = v$$

Since the applied voltage is time-variant, $v(t)$, the above equation can be written as,

$$Ri(t) + L\frac{di}{dt} + \frac{1}{C}\int_{-\infty}^{t} i(\tau)\, d\tau = v(t)$$

The differential equation format of the above equation becomes,

$$\frac{d^2 i}{dt^2} + \frac{R}{L}\frac{di}{dt} + \frac{1}{LC} i(t) = \frac{1}{L}\frac{dv}{dt}$$

This is a second order differential equation. Let us define two key parameters ζ and ω_0 (Both of which are measured as radians per second), as follows.

$$\zeta = \frac{R}{2L}$$

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

Substituting these parameters into the differential equation, we sometimes prefer to write the above equation in a standard form as follows.

$$\frac{d^2 i}{dt^2} + 2\zeta\frac{di}{dt} + \omega_0^2 i(t) = \frac{1}{L}\frac{dv}{dt}$$

3.6 BASIC MODELING ELEMENT IN FLUID SYSTEMS

Another important elemental system of MEMS is the fluid system. The three basic modeling elements in fluid systems are,

- Inertance
- Fluid resistance
- Fluid capacitance

Inertance, fluid resistance and fluid capacitance are similar to spring, damper and mass of the mechanical system.

3.6.1 Inertance

The governing law (Elemental equation) with respect to inertance is that the pressure momentum is the function of flow rate. The inertance is represented as

$$\Delta = \Gamma Q \tag{3.17}$$

where, Δ is pressure momentum, Γ is *inertance* and Q is flow rate. The element possessing inertance is known as *inertor*. The energy stored in the inertor is given by,

$$E = \frac{1}{2}\Gamma Q^2 \tag{3.18}$$

The value of inertance depends on the geometry of the inertor. An ideal inertor characterizes frictionless incompressible flow in a uniform passage. The fluid inertance of such ideal inertor is given by

$$\Gamma = \frac{\rho L}{A} \tag{3.19}$$

where, ρ is the density of the fluid, L is length of the inertor and A is the passage. Figure 3.8 describes the symbol of the fluid inertor.

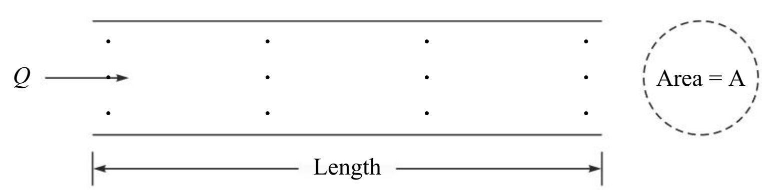


Fig. 3.8 *Fluid inertor*

3.6.2 Fluid Resistance

A fluid resistor dissipates energy. A flow of fluid through a fluid system gives rise to a pressure drop analogous to voltage drop across an electrical resistance element. The drop signifies loss. If the relationship between the fluid flow and pressure drop across such element is linear then the elemental equation can be written as,

$$Q = G_f P = \frac{1}{R_f} P \tag{3.20}$$

where, Q is fluid flow rate, P is the pressure difference, G_f and R_f are called fluid conductance (m^5/N-s) and fluid resistance (N-s/m^5) respectively. A fluid resistor characterized by its resistance is usually called the *fluid conductor*. The power dissipated through the element is governed by the equation,

$$\wp_f = PQ \tag{3.21}$$

The above model equation can be re-written in terms of fluid resistance as,

$$\wp_f = R_f Q^2 = \frac{1}{R_f} P^2 \tag{3.22}$$

3.6.3 Fluid Capacitor

An element in which the stored energy is a function of fluid pressure is defined as *fluid capacitor*. A fluid capacitor is defined by the following elemental equation.

$$V = C_f P \tag{3.23}$$

where, V is the volume (m^3) of the fluid in the capacitor, P is the fluid pressure, and C_f is constant called fluid capacitance (m^3/N). The above equation can be expressed in terms of flow rate.

$$\frac{dV}{dt} = C_f \frac{dP}{dt} \tag{3.24}$$

$$\Rightarrow \qquad Q = C_f \frac{dP}{dt}$$

The symbol of the ideal fluid capacitor is shown in Fig. 3.9, which is a kind of open tank filled with incompressible fluid. The capacitance of the fluid capacitor depends on the geometry of the tank, and is given by,

$$C_f = \frac{A}{\rho g} \tag{3.25}$$

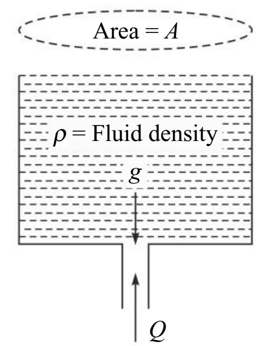


Fig. 3.9 *A fluid capacitor*

3.7 BASIC MODELING ELEMENTS IN THERMAL SYSTEMS

Thermal devices operate based on the principle of thermoelectric effect, thermoresistive effect and the pyroelectric effect. The devises are especially employed for driving, switching, and data reading/recording applications in communication, consumer products, vehicles, computing and in many other sectors. Although many MEMS systems include rotary as well as sliding systems for actuation purposes, thermal actuation is considerably preferred because of many novel features and characteristics the design possesses.

Regardless of the principle of operation of thermal device, a component in a MEMS may deform because of parasitic thermal intrusion probably due to current or friction. Thermal deformation is the increase of dimension of the MEMS part because of heat generation. The increase in dimension causes mid-way seizure of the system. Such parts might have a reputation of not being reliable over a period of time for which performance profile deteriorates. In order to deal with the problems associated with the design of such systems that are susceptible to thermal deformation, thermal characterization and modeling is essential. Considerable attention must be paid to encounter the thermal deformation problems. From the modeling and analysis, the behavior of the system is studied and the performance of the system can thus be predicted.

Thermal systems are characterized by heat and temperature. The associated energy within the system is referred to as thermal energy, which is stored, dissipated and transferred across the system. Heat is transferred by conduction, convection and radiation. The governing law for balancing the thermal energy is based on rate of energy flow into the system, rate of energy stored in the system and the rate of loss within the system. One fundamental law with regard to thermal system is that at absolute zero temperature the kinetic energy of the thermal system is zero. The basic modeling elements of thermal systems are

- Thermal capacitance
- Thermal resistance

No thermal inductance exists. A thermal system may be considered as a network of thermal resistances and capacitances linking different regions and representing conductive, convective, and radiative and heat storage processes.

3.7.1 Thermal Capacitance

Thermal capacitance is the basic parameter with regard to lumped model of the thermal system. The capacity of the thermal system to store the thermal energy, which is simply the heat, is reflected by the parameter thermal capacitance. The elemental equation is derived from the fact that the heat is a function of temperature. The following equation can be written

$$H = C_t \kappa \tag{3.26}$$

where, H is heat (Joules), κ is temperature and C_t, a constant is the thermal capacitance (Joule/K). Taking the derivative of the above equation,

$$\frac{dH}{dt} = C_t \frac{d\kappa}{dt} \qquad \Rightarrow \qquad C_t = \frac{q}{\tau} \tag{3.27}$$

The left hand side of the above equation is the rate of change of heat flow or rate of energy storage (q), τ is the rate of temperature. The equation signifies that if the thermal capacitance of the thermal system is large, the rate of temperature change due to heat is low.

3.7.2 Thermal Resistance

Pure thermal resistance neither characterizes loss nor stores any energy. The thermal resistance is defined by the following equation

$$R_t = \frac{\kappa_2 - \kappa_1}{q} \tag{3.28}$$

where, κ_1 and κ^2 are the different temperatures at two points, q is the flow rate and R_t is thermal resistance (K/W). The reciprocal of thermal resistance is called heat conductance (W/K). Thermal resistances are distinguishably different for the three different mechanisms of heat transfer such as conduction, convection and radiation.

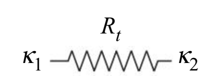


Fig. 3.10 *Thermal resistance*

A transfer of heat energy occurs when systems having different temperatures are allowed to mix. By this process the total amount of energy remains constant. The quantity of heat given off is equal to the quantity of heat energy gained. The transfer of energy will continue in this way until both systems reach the same temperature. The flow can be controlled by suitably establishing the system parameters and variables, which is optimized through modeling.

3.8 SUMMARY

All engineering systems are classified as mechanical system, electrical system, fluid system or thermal system. Basic modeling elements are *pure*. Table 3.3 summarizes the modeling elements of the pure system you have already studied. A system could either be a pure mechanical system, pure electrical system, pure fluid system, pure thermal system or any combinations of these (hybrid). A real practical system may either contain more pure elements from different systems. Such systems are called hybrid systems. Mostly, MEMS systems are hybrid types. In order to model the MEMS system, fundamental knowledge as to how the hybrid systems are built up plays important role. Next section discusses with the build up process in relation to the modeling of pure and hybrid systems.

Table 3.3 *Basic modeling elements of the systems*

Engineering systems	*Energy source (Means of supplying energy)*	*Basic modeling elements*	*Energy stored/Power loss*
Mechanical	• Force	Translational spring	Translational = $\frac{1}{2}kx^2$
	• Velocity	Rotational spring	Rotational = $\frac{1}{2k_a}T^2$
		Translational damper	Power loss = $b_i v^2$
		Rotational damper	Power loss = $b_r w^2$
		Mass/inertia	Translational = $\frac{1}{2}mv^2/$
			Rotational = $\frac{1}{2}Jw^2$
Electrical	• Current	Inductor	$\frac{1}{2}Li^2$
	• Voltage	Resistance	Power loss = $\frac{v^2}{R}$
		Capacitance	$\frac{1}{2}Cv^2$
Fluid	• Fluid flow	Inertance	$\frac{1}{2}\Gamma Q^2$
	• Pressure	Fluid resistance	$\frac{1}{R_f}P^2$
		Fluid capacitance	$\frac{1}{2C_f}P^2$
Thermal	• Heat flow	Thermal capacitance	$C_t \kappa$
	• Temperature	Thermal resistance	

3.9 TRANSLATIONAL PURE MECHANICAL SYSTEM WITH SPRING, DAMPER AND MASS

If a system inherits many similar elements they are tied together. This means all spring elements are combined into one, all dampers are combined to a single damper and so on in order to describe the system entirely.

In general, engineering systems are modeled by following some principle. Having identified the basic elements in the system, the balancing laws are utilized in order to derive the input-output model equations. The laws of forces or torque (Newton's laws or D'Alembert's law) and the law of displacement or rotation are principally applied. In essence, the sum total of all the forces or torques acting on the system has to be balanced with effective force or torque and the sum total of all the linear or angular displacements has to be equated to zero.

$$\sum_k (F)_k = m\frac{dv}{dt} \quad \text{or} \quad \sum_k (T)_k = J\frac{dw}{dt} \tag{3.29}$$

$$\sum_k (\Delta x)_k = 0 \quad \text{or} \quad \sum_k (\Delta \theta)_k = 0 \tag{3.30}$$

where,

F_i = The *ith* force acting on the system
m = Mass
v = Velocity of the translational system
dv/dt = Acceleration of the translational system
Δx_i = Translational displacement due to *ith* force
T_i = The *ith* torque acting on the system
J = Inertia
w = Angular velocity of the rotational system
dw/dt = Angular acceleration of the system
$\Delta\theta_i$ = Angular displacement due to *ith* torque

Figure 3.11(a) illustrates a translational mechanical system described by spring, damper and mass element. It is the schematic representation of typical translational mechanical systems inheriting all the three basic elements. Figure 3.11(b) shows another mechanical system with only spring and mass element. Here the loss within the system has been assumed to be zero and therefore neglected. Such a system is called undamped system because if force is applied the system would oscillate as the loss is zero.

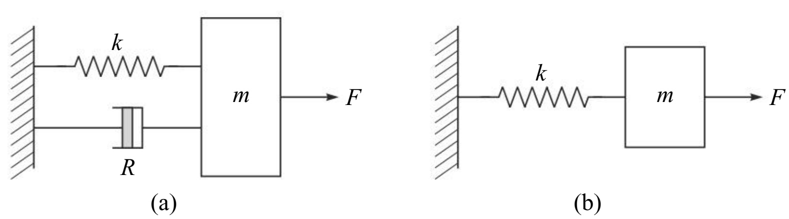


Fig. 3.11 *(a) Translational mechanical system with spring, damper and mass, (b) A pure mechanical system without damper*

When a force is applied to the mechanical system shown in the Fig. 3.11(a), energy is stored in the spring as well as in the mass elements. Energy/power is dissipated/lost in the damper or dashpot element. If multiple forces act on the system the resultant force has to be calculated. As before, the energy balance equation can be developed as follows.

Let,

F = Applied force; x = Resulting displacement; $v = \dfrac{dx}{dt}$ = resulting velocity; $a = \dfrac{dv}{dt} = \dfrac{d^2x}{dt^2}$ = resulting acceleration; m = Mass of the mechanical system; k = Translational spring constant or stiffness; and b_t = Translational damping constant;

The potential energy stored in the spring is,

$$E = \frac{1}{2}kx^2 \tag{3.31}$$

The equivalent force required to store the above potential energy is,

$$F_p = kx \tag{3.32}$$

The kinetic energy stored in the mass,

$$E_k = \frac{1}{2}mv^2 \tag{3.33}$$

The equivalent force required to store the above kinetic energy is,

$$E_k = ma = m\frac{d^2x}{dt^2} \tag{3.34}$$

The loss in the damper is,

$$b_t v^2 \tag{3.35}$$

The equivalent force (called damping force) required for the above loss is,

$$F_l = b_t v = b_t \frac{dx}{dt} \tag{3.36}$$

At any point of time the sum total of the energy is balanced by the elements. Alternatively, the sum total of equivalent forces is equal to the force applied to the system. That is,

$$F = F_p + F_k + F_l \tag{3.37}$$

Substituting Eqs. 3.32, 3.34 and 3.36 in Eq. 3.37,

$$F = m\frac{d^2x}{dt^2} + b_t\frac{dx}{dt} + kx \tag{3.38}$$

The above equation is called the *model equation* of the mechanical system described in Fig. 3.11(a). The model equation of the spring-mass type mechanical system, shown in Fig. 3.11(b), with applied force zero, can be written as,

$$m\frac{d^2x}{dt^2} + kx = 0 \tag{3.39}$$

The above two mechanical systems are called second order systems because of the appearance of the second derivative term, i.e. d^2/dt^2.

3.10 ROTATIONAL PURE MECHANICAL SYSTEM WITH SPRING, DAMPER AND MASS

Figure 3.12 illustrates a rotational mechanical system with spring, damper and mass. This is the schematic representation of many kinds of rotational mechanical systems inheriting all the basic elements.

Let T = The applied torque; θ = The resulting angular displacement; $w = \frac{d\theta}{dt}$ = resulting velocity; $\alpha = \frac{dw}{dt} = \frac{d^2\theta}{dt^2}$ = resulting acceleration; J = Inertia of the mechanical system; k = rotational spring constant or stiffness; and b_t = rotational damping constant.

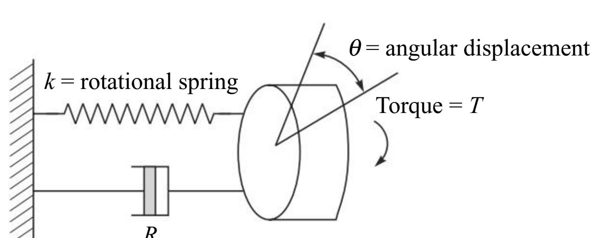


Fig. 3.12 *Rotational mechanical system with spring, damper and mass element*

By similar approach the model equation of the rotational mechanical system can be written as

$$T = J\frac{d^2\theta}{dt^2} + b_t\frac{d\theta}{dt} + k\theta \tag{3.40}$$

3.11 MODELING HYBRID SYSTEMS: AN EXAMPLE WITH PARALLEL PLATE MEMS VARACTOR

Varactor is a variable capacitor in which the value of capacitance can be varied by some means, usually by a voltage. Varactors are mainly used for resonance or tuning applications. Tuning is a process of selecting a single desired frequency component or a group of frequency components. Consider the following LC circuit (refer to Fig. 3.13). In Fig. 3.13(a) the frequency produced by the circuit is given by,

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{3.41}$$

In Fig. 3.13(b) a varactor is connected to the previous LC circuit, in parallel. Let ΔC be the maximum change in the capacitance with respect to the nominal capacitance C. Then the minimum frequency will be,

$$\begin{aligned} f_{\min} &= \frac{1}{2\pi\sqrt{L(C+\Delta C)}} = \frac{1}{2\pi\sqrt{LC}\,(1+\Delta C/C)^{1/2}} \\ &= f_0\left(1+\frac{\Delta C}{C}\right)^{-\frac{1}{2}} \approx f_0\left(1-\frac{\Delta C}{C}\right) \end{aligned} \tag{3.42}$$

Now, the range of frequency of operation will be $f_{\min}$ to f_0. The above equation states that the frequency of operation depends on the value of change in capacitance, ΔC. That is if the value of capacitance changes, then the frequency of operation will change accordingly. Alternatively, whenever there is a requirement to vary the range of frequency of operation, varactor is the obvious choice. Traditional semiconductor junction diode based varactors have long been in use. MEMS based varactor is a recent development. MEMS varactor is simply a parallel plate capacitor in which the position of one plate is changed to control the capacitance from the nominal value.

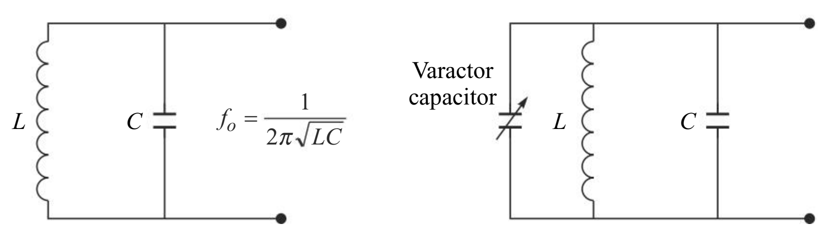


Fig. 3.13 *(a) An LC circuit, (b) An LC tuning circuit*

As already stated, varying the voltage across the parallel plates can change the value of the capacitance in a varactor. Refer to Fig. 3.14. Now let us define V_i and C_i as the instantaneous values of the applied DC voltage and corresponding capacitance (which ranges from C to $(C + \Delta C)$) in the circuit. The applied DC voltage is called tuning voltage. The instantaneous value of capacitance is given by,

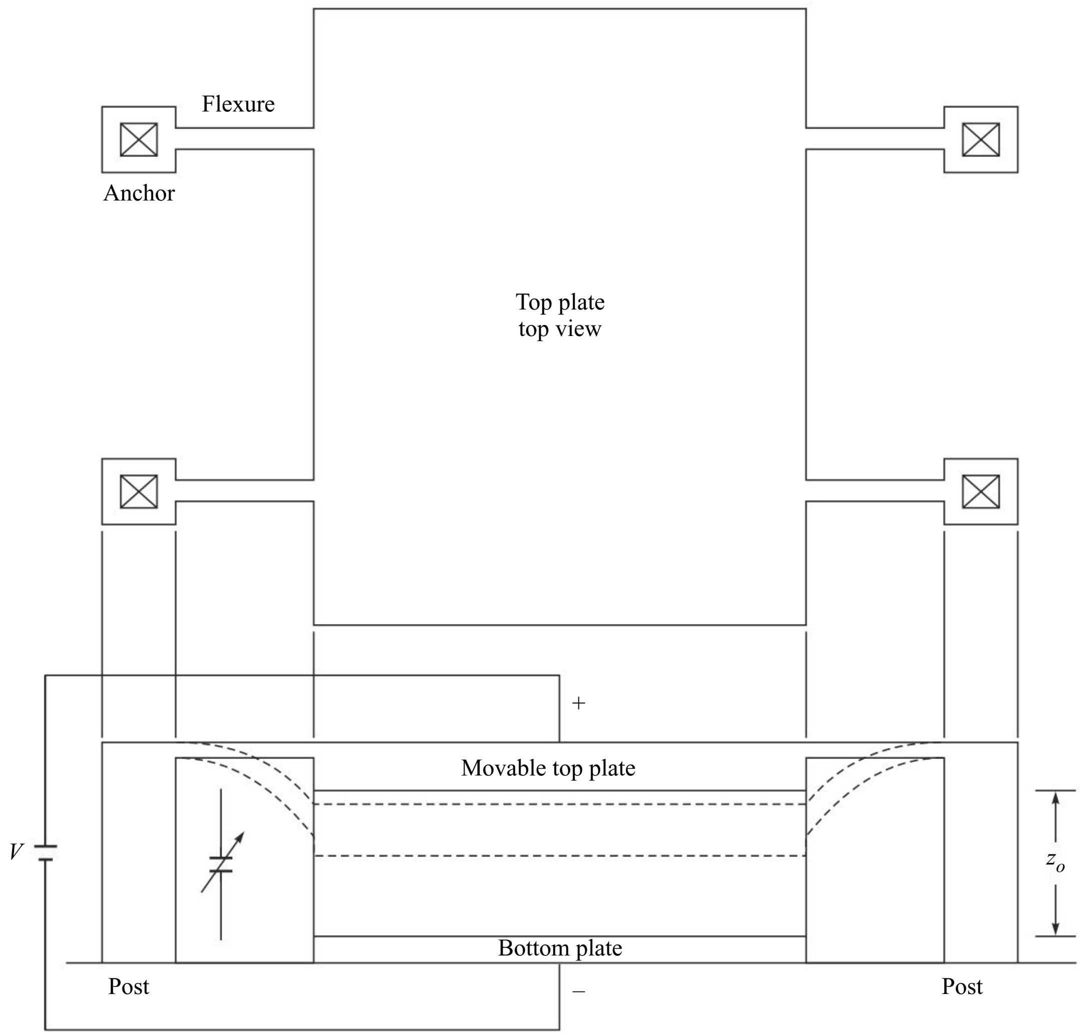


Fig. 3.14 *MEMS Varactor*

$$C_i = \frac{\varepsilon A}{z_0 - z} \tag{3.43}$$

$$\Rightarrow \qquad \varepsilon A = C_i (z_0 - z)$$

where, ε is the permittivity of the air, A is the effective plate overlap area and $(z_0 - z)$ is the change in top-plate position (z_0 is nominal distance between the plates when input voltage is zero). As you know that when a potential difference is applied between the two parallel plates, an electric filed is established. The relationship between the applied voltage and the electric field, E_i is given by,

$$V_i = E_i (z_0 - z) \tag{3.44}$$

$$\Rightarrow \qquad E_i = \frac{V_i}{(z_0 - z)}$$

The electric field causes the parallel plates to attract each other. The force of attraction is called electrostatic force, denoted by F_e. Since one plate is fixed (bottom plate), the other will be displaced and the separation between them will be lesser. The displacement is proportional to the applied voltage. If Q is the charge in each plate the magnitude of the force on each plate is given by,

$$F_e = \frac{QE_i}{2} \tag{3.45}$$

where, $$Q = C_i V_i \tag{3.46}$$

By putting the value of E_i from Eq. 3.44 and the value of Q from Eq. 3.46, Eq. 3.46 can therefore be written as,

$$F_e = \frac{\varepsilon A V_i^2}{2(z_0 - z)^2} \tag{3.47}$$

Putting the value of εA from Eq. 3.43 the above equation can further be written as,

$$F_e = \frac{C_i V_i^2}{2(z_0 - z)} \tag{3.48}$$

The electromechanical dynamic behavior of the varactor is described by a second order system accommodating the constituent force components such as inertia, viscous damping force and spring force. At equilibrium, the general mathematical expression is given by,

$$m\frac{d^2z}{dz^2} + b\frac{dz}{dz} + F_m = F_e \tag{3.49}$$

where, m is the mass of the top plate, b is the damping constant and F_m is the spring force. If the inertia and damping forces are assumed to be zero, then the above equation can be reduced to,

$$F_m = F_e \tag{3.50}$$

Since, F_m equals to $k_m z$, where k_m is the mechanical spring constant associated with the plate-anchor arrangement, the above equation can be written as,

$$k_m z = \frac{C_i V_i^2}{2(z_0 - z)} \tag{3.51}$$

From the above, the *DC* tuning voltage can be expressed as,

$$V_i = \sqrt{\frac{2k_m z (z_0 - z)^2}{\varepsilon A}} \tag{3.52}$$

Let us define k_e as the effective spring constant with respect to electrostatic force. Its mathematical representation is given by,

$$k_e = \frac{\partial F_e}{\partial z} \tag{3.53}$$

It can further be expanded as,

$$k_e = \frac{C_i V_i^2}{(z_0 - z)^2} \tag{3.54}$$

The effective spring constant can be expressed in terms of k_m as follows.

$$k_e = \frac{2k_m z}{(z_0 - z)} \tag{3.55}$$

3.12 ANALOGY BETWEEN 2nd ORDER MECHANICAL AND ELECTRICAL SYSTEMS

The systems are modeled by virtue of governing laws. One of the ways the systems are modeled is based on the input–output relationships. For example, for a given input voltage, $v(t)$ to a resistor element (electrical system) the resulting output current, $i(t)$ is adequately governed by the fundamental laws of electricity, which in turn is accurately represented by input-output model equation as given below.

$$i(t) = G \times v(t) \tag{3.56}$$

where, $v(t)$ is the input voltage, $i(t)$ is the current and G is a constant that physically characterizes the system, i.e. depends on the nature of the system. For this reason the constant is called *physical parameter*. The above equation can be written in a standard form as shown below.

$$a_1 y(t) = b_0 u(t) \text{ or simply } a_1 y = b_0 u \tag{3.57}$$

where, y represents the output, u is the input, the coefficients a_1 and b_0 are constants characterizing the system. The above input-output model equation is a liner one that does not contain a derivative term. However, mostly, the input–output model equations of dynamic system contain derivative terms constituting linear ordinary differential equations. For instance, the input-output model equation of a thermal system is given by

$$RC \frac{d\kappa}{dt} + \kappa = \kappa_0 \tag{3.58}$$

where, $d\kappa/dt$ is the rate of change of temperature, κ_0 is the temperature difference between the two points of interest, R is called thermal resistance and C is called the thermal capacitance. As an example, consider a thermometer that contains a bulb of mercury. The driving force for the change of the effect is the difference between the temperature of the human body and the temperature of the thermometer. This driving force decreases to zero as the thermometer heats up. Equation 3.58 can be written in a standard form as given below

$$a_1 \dot{y} + a_0 y = b_0 u(t) \tag{3.59}$$

where, y and u represent output and input of the system, respectively. $\dot{y}$ is the first derivative of output function y. The coefficients a_1, a_0 and b_0 are the constants characterizing the system.

Similarly, by applying fundamental laws to the hybrid type systems such as electromechanical systems (dc motor, for example), the input-output model equation can directly be written as,

$$J \frac{d^2\theta}{dt^2} + R \frac{d\theta}{dt} + k\theta = K_t I_a \tag{3.60}$$

where, J is the moment of inertia of the rotor, R is the rotational damping constant, k is the rotational spring constant or stiffness; k_t is the torque constant; and I_a is the armature current. Equation 3.60 can also be written in the standard form as follows

$$a_2 \ddot{y} + a_1 \dot{y} + a_0 y = b_0 u(t) \tag{3.61}$$

As before, the coefficients a_2 a_1, a_0 and b_0 are the constants characterizing the system. $\ddot{y}$ and $\dot{y}$ are the second and first derivatives of the output function, respectively. In Eq. 3.56, input is the voltage, output is the current; in Eq. 3.58, the input is the difference in temperature, output is the rise in temperature (i.e. the temperature change). On the other hand, in Eq. 3.60 the input is the armature current, (I_a) and the resulting output is the angular displacement (θ). System representing Eq. 3.56 bears a linear relationship between the input and output and it does not utilize ordinary differential equation. On the other hand, Eq.

3.58 and 3.60 utilize ordinary differential equations in order to describe the system's input-output relationships.

So, a typical second order mechanical system with an input forcing function $f(t)$ is generally described by

$$m\frac{d^2y}{dt^2} + b\frac{dy}{dt} + ky = f(t) \tag{3.62}$$

Similarly, the second order electrical system with a forcing function $e(t)$ is described by

$$L\frac{d^2i}{dt^2} + R_\Omega\frac{di}{dt} + \frac{1}{C}i = \frac{de(t)}{dt} \tag{3.63}$$

where, m = mass, b = damping constant, k = spring constant, $f(t)$ = input force, y = output displacement, L = inductance, R_Ω = resistance, C = capacitance, $e(t)$ = input voltage, and i = output current.

From the above two equations, analogy can be made. Refer to Table 3.4.

3.13 PROPERTIES OF MATERIALS

Study of properties of material plays an important role in MEMS design. Development of new materials for microstructure applications is underway. New materials are being developed taking into account physical, mechanical, electrical and thermal performances. In order to quantify the performance it is first necessary to study the quantifying parameters. This in turn entails the study of properties of materials. Properties of common solid materials are divided into four categories.

Table 3.4 *Analogy between the elemental mechanical and electrical systems*

Mechanical Systems	Electrical Systems
Force	Rate of change of electric potential
Spring	Reciprocal of capacitor
Damper	Resistor
Mass/Inertia	Inductor

- Physical
- Mechanical
- Thermal
- Electrical

The important physical parameters, which describe the physical property of materials, are density and melting point. Mechanical properties are described by elastic modulus, shear modulus, Poisson's ratio, yielding stress, ultimate stress and elongation. Similarly, thermal properties are characterized by thermal coefficient of expansion, heat capacity, and thermal conductivity. Finally, resistivity describes the electrical properties of materials. The above properties and their expressive parameters are presented in Table 3.5. The very precise definitions of these parameters follow. Table 3.6 describes physical, mechanical, thermal and electrical properties of some materials. Table App-5 in Appendix B shows some important properties of materials.

- ***Density:*** Density is defined as the mass per unit volume of a material.
- ***Hardness:*** The hardness of a material measures how tightly the atoms are held together within it.
- ***Porosity:*** A measure of ability to allow the passage of a air or fluid.
- ***Elongation:*** The elongation is the percentage increase in length that occurs before it breaks under tension.
- ***Elastic modulus:*** The elastic modulus is described as the relative stiffness of the material within the elastic range. It is determined from a stress–strain curve in terms of calculating the ratio of stress (force per unit area) to strain (deformation) under tension or compression.

Table 3.5 *Characterizing properties of materials*

Sl No.	*Properties*	*Parameters/Constant*
1	Physical	Density Hardness Porosity
2	Mechanical (basic)	Elastic modulus Shear modulus Poisson's ratio
	Mechanical (strength)	Yielding stress Ultimate stress Elongation
3	Thermal	Coefficient of thermal expansion Thermal conductivity Heat capacity Melting point
4	Electrical	Electric resistivity Dielectric constant Work function

- ***Poisson's ratio:*** Poisson's ratio is the ratio of transverse contraction strain to longitudinal extension strain in the direction of stretching force.
- ***Shear modulus:*** The ratio of shearing stress to the shearing strain within the proportional limit of a material is called shear modulus.
- ***Yielding stress:*** The yield stress corresponds to a point where the material begins to have unrecoverable deformation. The stress and strain corresponding to the yield point are called the *yield strain*.
- ***Ultimate stress:*** The maximum tensile stress a material is capable of carrying is known as the *Ultimate tensile stress.* It corresponds to the highest point on the tensile stress–strain curve. The corresponding strain is known as the *Ultimate tensile strain.*
- ***Coefficient of thermal expansion:*** The coefficient of thermal expansion is defined as the fractional increase in length per unit rise in temperature.
- ***Thermal conductivity:*** Thermal conductivity expresses the heat flux that will flow if a certain temperature gradient exists over the material. It is expressed in W/m.K.
- ***Heat capacity:*** The amount of energy required to raise the temperature of a substance by 1 °K is the heat capacity. If the substance has a unit mass, the amount is referred to as specific heat capacity, or specific heat.
- ***Melting point:*** The melting point of a solid is the temperature at which it changes state from solid to liquid.
- ***Dielectric constant:*** The dielectric constant of a medium is its ability to reduce the force F of attraction of charged (q_1 and q_2) particles separated at distance r, compared to vacuum.
- ***Electric resistivity:*** The constant of proportionality between the potential difference and the resulted current is called the *resistance*. Resistance is also proportional to the length and inversely proportional to the cross-sectional area of the material. Therefore, resistivity (Ohm-meter) is defined as the resistance per unit length times cross-sectional area.

Table 3.6 *Properties of material*

Material	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Aluminium (Al)	20	70	60	70	26	0.33	23.0	237	2.71	6660.3	2519	27
Aluminium Alloy	35–500	100–500	1–45	70–79	26–30	0.33	23.0	-	2.64–2.8	565.0–660.0	2519	50
Brass	70–550	200–620	4–60	96–110	36–41	0.34	19.1–21.2	-	8.4–8.75	930.0	-	20–61
Brass; Noval	170–410	410–590	15–50	100	39	0.34	21.1	-	8.4	-	-	
Brass; Red (80%Cu, 20%Zn)	90–470	300–590	4–50	100	39	0.34	19.1	-	8.75	1000	-	
Bronze; Regular	82–690	200–830	5–60	96–120	36–44	0.34	18.0–21.0	-	7.8–8.8	1050	-	
Bronze; Manganese	170–450	450–620	10–35	100	39	0.34	20.0	-	8.3	-	-	
Copper (Cu)	55–330	230–380	10–50	110–120	40–47	0.33–0.36	16.6–17.6	410	8.94	1085	2562	17
Copper Alloy	760	830	4	120	47	-	17.0	-	8.23	925.0	-	17–490
Glass	-	30–1000	-	48–83	19–34	0.2–0.27	5.00–11.0	-	2.4–2.8	-	-	
Iron (Cast)	120–290	69–480	0–1	83–170	32–69	0.2–0.3	9.90–12.0	-	7–7.4	-	-	
Iron (Wrought)	210	340	35	190	75	0.3	12.0	-	7.4–7.8	-	-	
Magnesium (Mg)	20–70	100–170	5–15	41	15	0.35	25.2	156	1.74	650.0	1090	45
Magnesium Alloy	80–280	140–340	2–20	45	17	0.35	26.1–28.8	-	1.77	1246	2061	
Monel (67%Ni,30%Cu)	170–1100	450–1200	2–50	170	66	0.32	14.0	-	8.84	1330	-	
Nickel(Ni)	140–620	310–760	2–50	210	80	0.31	13.0	90.7	8.89	1455	2913	70
Nylon;Polyamide	-	40–70	50	2.1–2.8	-	0.4	75.0–100	-	1.1	-	-	
Solder;Tin-Lead	-	12–54	55–30									130–210
Steel	280–1600	340–1900	3–40	190–210	75–80	0.27–0.3	10.0–18.0	-				120–1700
Titanium (Ti)	-	500	25	110	40–40	0.33	-	21.9				430
Titanium alloy	-	900–970	10	110–120	39–44	0.33	8.00–10.0	-				
Tungsten (W)	-	1400–4K	0–4	-			4.30	174				56

A = Yield Stress (MPa); B = Ultimate Stress (MPa); C = Elongation (%); D = Elastic Modulus (GPa); E = Shear Modulus (GPa); F = Poisson's Ratio; G = Thermal Expansion Coefficient ($\times 10^{-6}/°C$); H = Thermal Conductivity (W/m.K); I = Density ($\times 1000\ kg/m^3$); J = Melting Point (°C); K = Boiling Point (°C) and L = Electric Resistivity ($\times 10^{-9}\Omega \cdot m$)

- ***Work function:*** The minimum energy, eV (elecronvolt) required to extract an electron from a solid. It signifies how tightly electrons are bound to a material (Refer Table 3.7).

3.14 RELATIONSHIP BETWEEN YOUNG'S MODULUS (*E*), BULK MODULUS (K), SHEAR MODULUS (*G*) AND POISSON'S RATIO

The Young Modulus, E is one of the most important properties of material. The material property that describes its stiffness is referred to as Young's modulus. The modulus is measured by pulling a sample of a material in a tensile testing machine, an instrument that measures force. The term, ε called strain and is defined in elementary form as the ratio of change in length to the original length. That is, $\varepsilon = \Delta L/L$

Table 3.7 *Work function for some materials*

Materials	*Work function (eV)*
Sodium	2.75
Silicon (111)	4.6
Silicon (100)	4.85
Graphite	5.0
Gold (111)	5.31
Gold (110)	5.37
Gold (100)	5.47

The Bulk Modulus is a measure of the change in volume in response to applied pressure. Bulk modulus is thus an index which provides information with regard to the change in volume of a solid substance as the pressure on it is changed. Bulk modulus is also called elastic modulus. Mathematically it is expressed as follows.

$$K = -V\frac{\partial P}{\partial V}$$

The above equation can also be written as,

$$K = -\rho\frac{\partial P}{\partial \rho}$$

where, P is pressure and V is volume, ρ is density. Typical value for the bulk modulus for steel is $K_s = 160 \times 10^9$ N/m^2. The reciprocal of the bulk modulus is defined as compressibility. Note that the amount of compression of solids and liquids is observed to be very small.

When a material is placed under a tensile stress, an accompanying strain is developed in the same direction. As a consequence of this elongation, there should be constrictions in the other two directions. The dimensional changes are expressed by two ratios called lateral to axial strains, respectively. The ratio of these ratios is further expressed as Poisson's ratio (ξ). Thus, Poisson's ratio is the ratio of the lateral (compressive) to axial (tensile) strains. In other words, Poisson's ratio $\boldsymbol{v}$ is the ratio of transverse contraction strain to longitudinal extension strain in the direction of stretching force. The lateral and axial deformations are taken as negative and positive, respectively. The definition of Poisson's ratio itself already contains a minus sign as expressed below, such that the materials can have a positive value.

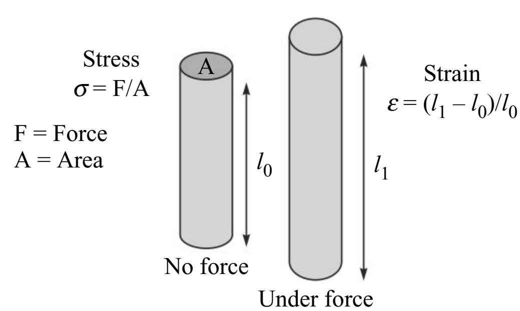


Fig. 3.15 *Illustration that shows Stress, Strain, and the Young's modulus*

$$\xi = -\varepsilon_{\text{trans}}/\varepsilon_{\text{longitudinal}}$$

The ratio of shearing stress τ to shearing strain γ within the proportional limit of a material is called shear modulus. It is sometimes also called the rigidity. Mathematically it is expressed as,

$$G = \frac{\text{shear stress}}{\text{shear strain}} = \frac{\tau}{\gamma}$$

There exist relationships between the modulus defined above. Poisson's ratio is related to Young's modulus E, elastic moduli K, and shear modulus G by the following.

$$E = 2G(1 + v)$$
$$v = (3K - 2G)/(6K + 2G)$$

3.15 SUMMARY

This chapter described the need for modeling a system and listed the analysis and design procedure. The chapter begins with summarizing the design and development procedure. The need for system modeling has been emphasized with examples. It was necessary to define what a system means. Various types of systems, such as SISO, SIMO, MISO, and MIMO have been listed and discussed. The other form of classification in terms of elemental systems, taking into account the law of conservation of energy, such as mechanical, electrical, fluid and thermal system has been presented. Basic modeling elements of the each elemental system are discussed and a summary is presented in a tabular form. Examples are given in order to enhance the understanding of system modeling. Typical model examples such as translational and rotational mechanical system containing all the basic elements such as spring, damper and mass/inertia are considered in this respect. The analogy between the mechanical system and electrical system has been drawn from their respective model equations. Finally the properties of materials are discussed. Important physical, mechanical, electrical and thermal properties of the materials are defined and presented in tabular form.

Points to Remember

- System modeling is about solving practical problems by the use of mathematical models, called model equations, which are formulated looking at the real physical system.
- The results of the analysis help how well the system met the design parameters and how effectively to alter them to get best design requirement.
- A *system* is defined as a group of properly arranged *elements* those when act together provides the *desired* output with respect to *available* inputs.
- Mechatronic systems are not elemental system rather they are hybrid type.
- The noise energy is the *unwanted available input energy*.
- Distinguishably three *basic modeling elements* are earmarked for the mechanical systems.
- Spring element stores potential energy.
- The reciprocal of the stiffness is called mechanical capacitance or *compliance*.
- Damper element does not store any energy. It consumes energy, which cannot be recovered. Other name of the damper is *dashpot* that symbolizes resistance, more appropriately, *mechanical resistance*.
- System model based on the elements such as spring, damper and mass/inertia is referred to as *lumped-parameter* model due to the reason that each element is basic and independent of the other.
- The element possessing inertance is known as *inertor*.
- A fluid resistor dissipates energy.

- An element in which the stored energy is a function of fluid pressure is defined as *fluid capacitor*.
- Thermal systems are characterized by heat and temperature.
- Thermal deformation is the increase of dimension of the mechanical part because of heat generation.
- Heat is transferred by conduction, convection and radiation.
- One fundamental law with regard to thermal system is that at absolute zero temperature, the kinetic energy of the thermal system is zero.
- The capacity of the thermal system to store the thermal energy is reflected by the thermal capacitance.
- Pure thermal resistance neither characterizes loss nor stores any energy.
- When the system parameters change with respect to time it is called dynamic system.
- New materials are being developed taking into account their physical, mechanical, electrical and thermal performances.
- Properties of common solid materials are divided into four categories such as physical, mechanical, thermal and electrical.
- The important parameters, which describe the physical property of materials, are density and melting point.
- Mechanical properties are described by elastic modulus, shear modulus, Poisson's ratio, yielding stress, ultimate stress and elongation.
- Thermal properties are characterized by thermal coefficient of expansion, heat capacity, and thermal conductivity.
- Resistivity describes the electrical properties of materials.

Exercises

1. Define the 'system' from the viewpoint of engineering study. Distinguish between the macro system and the micro system. Regardless of their physical dimensions, how can you categorize them with respect to number of inputs and outputs? Is their any other form of categorization?
2. What do you mean by system modeling? What intended benefits are achieved through modeling? Give an example of modeling a typical macro system. How does simulation relate to modeling?
3. What is the suggested procedure a designer should follow while designing a system?
4. Represent a system in a block diagram and show various types of available inputs and outputs. Distinguish between noise and faults.
5. What do you mean by elemental system? Mention the four types of elemental systems and their respective energy sources.
6. Define the term spring, damper and mass of an elemental mechanical system. Write down the elemental modeling equations of these elements respectively and define the terms used therein. Also distinguish between the translational and rotational notations, if any, used in the equations of the elemental mechanical systems.
7. What are the basis elements of an elemental electrical system? Write down the elemental modeling equations of these elements and define the terms used. Make an analogy between the elements of mechanical and electrical systems. Mention the element that stores kinetic energy in case of electrical system?
8. The fluid system is also described by the basic modeling elements. What are they? Write down the elemental modeling equations of these elements and define the terms used. With regard to fluid capacitor element, show how gravity plays an important role in modeling the fluid system.

9. Write down the relationships between heat and temperature. What are the three methods by which the heat energy is transferred from one point to another?
10. How many basic elements exist in case of a thermal system? Define thermal capacitance and thermal resistance. Write down the elemental modeling equations of these elements and define the terms used therein. Draw the symbols for these elements.
11. Prepare a table summarizing the following.
    (a) Elemental systems
    (b) Their energy sources
    (c) Their basic modeling elements
    (d) Energy stored/Power lost
12. How can you represent a system, which consists of many elements of the same types? What principle is to be followed in such situations?
13. Represent a translational mechanical system, which has spring, damper and mass elements. Based on the law of conservation of energy and starting from the basic elemental equations derive the final model equation of this three-element complete mechanical system.
14. Represent a rotational mechanical system, which has rotational spring, rotational damper and inertia elements. Based on the law of conservation of energy and starting from the basic elemental equations, derive the final model equation of this three-element rotational mechanical system.
15. From their model equations, make an analogy between a mechanical system and an electrical system.
16. There are primarily four categories as far as properties of materials are concerned—name them. Define all the property parameters you know.
17. Describe the relationship between Young's modulus (E), Bulk modulus (K), Shear modulus (G) and Poisson's ratio.

Chapter 4

Passive Components and Systems

Objectives

The objective of this chapter is to study the following.

- Passive systems and their classifications
- The concept of System-on-a-Chip (SOC)
- Passive Electronic Systems (PES) in terms of discussion of various types of signal conditioning circuits such as clipping circuit, positive and negative clamper, multi-stage amplifier, instrumentation amplifier, comparator, bridge circuits, analog to digital (AD) conversion process and vice versa
- Passive Mechanical Systems (PMS) such as bearing, gears, rack and pinion, slider-crank, Geneva wheel, four-bar linkages, flexure and anchor

4.1 INTRODUCTION

From the viewpoint of integrated system, physically, a full-fledged MEMS device is composed of two main groups of functional systems: active and passive systems. Sensors and actuators are mostly considered as active systems. An active system could be a single component or composed of several components. Active components (described later) are the important functional building blocks around which passive components are interfaced in order to take electronic or mechanical advantage. A transducer is an active element in a sensor that produces a measurable response to a change in a physical condition such as temperature, pressure, humidity, flow, light intensity, magnetic field, vibration, and so on. It responds to some properties of the environment and directly transforms the response into an electric signal by adopting one of the transduction principle such as thermoelectric, photoelectric, electromagnetic, magnetoelectric, thermoelastic, pyroelectric and thermomagnetic. Actuator is an active device that makes something move, usually in a controlled manner. The controlled movement is called actuation, which is achieved through piezoelectric, hydraulic, electrostatic, magnetostrictive and thermal method.

Passive systems are those which when interfaced with the active system constitute a full fledged MEMS device. Passive elements are the components, which typically are not directly involved in

transducing power, rather they help in handling, manipulating and transferring power to a suitable form so that it will be easier to deal with. Passive systems are of two types; Passive Electronic Systems (PES) and Passive Mechanical Systems (PMS). PESs are amplifiers, filters, data converters, isolators, clippers, bridge circuits, and so on. PMS are gear, bearing, flexure, hinges, anchor, crank and slider, rack and pinion, and so on. The passive elements are responsible for transferring power from one place to another by virtue of what is known as 'electronic and/or mechanical advantage'. It is obvious that both PESs and PMSs facilitate electronic and mechanical advantage, respectively.

The role of PES in an integrated system is to optimally modify some raw or available signal to a form that will be suitable to use. One of the major applications are found in the MEMS sensors, although some MEMS actuators do not exclude them. Regardless of their types the PES are interfaced at the output of the transducer in order to provide refined output which can be used conveniently. A temperature, pressure, humidity, magnetic field transducer has an analog output that sometimes falls in the order of a few microvolts (if the output is a voltage signal). The level of transduced equivalence is so low that it essentially requires amplification in order to attain a desirable level so that it can be handled properly for subsequent use; may be at a control point. Eventually, the PES *conditions* the available output signal of the transducer. Similarly, with respect to mechanical systems, gears make it possible to change the rate of rotation of a shaft. They can even change the direction condition of the axis of rotation and can change rotary motion condition to linear one. Passive mechanical systems are called basic machines as they can facilitate *mechanism*. This chapter deals with the PES and PMS, which are auxiliary building blocks in MEMS.

4.2 SYSTEM-ON-A-CHIP (SoC)

The MEMS technology is moving forward to adopt the concept of system-on-a-chip (SoC) technology, which is defined as the integration and packaging of all the necessary electronic circuits and mechanical parts on a single integrated chip. A simple example is given in Fig. 4.1.

A system describing itself as a microelectromechanical system, could be anything, starting from a single but complex 3D microstructure to a multiple multi-layer super complex structure. Precisely, a system can be a microrobot, a complex array of microactuators, a phone, a camera, and so on. But a SoC means much more than what we think. SoC video and audio devices might be embedded in the brains and ear of blind deaf people allowing them to see and hear, respectively. Handheld computers with small whip antennas might be capable of browsing the Internet at megabit-per-second speeds. These are the forecasting news as far as futuristic SoC technology is concerned. Whatever may be the expectations the fundamental however, lies with the basic understanding of the individual components of the SoC technology. In the subsequent chapters we will concentrate on discussing the very basic and fundamental building blocks of MEMS sensors and actuators and their principle of operations. Besides sensors and actuators, the SoC technology, as expected, embeds electronic circuitry and interfacings in order to make the entire thing a fully integrated device. It appears that the necessary integrated electronic circuitry and other micromachined elements can be built simultaneously to form a complete system on a chip, which when mass-produced dramatically reduce the price with increased reliability. The electronic circuitry that needs to be integrated are amplifier, noise filter, bridge circuits, and so on. Figure 4.1(c) shows the layout of an SoC gas sensor with an array of four gas sensing elements, decoders, ADC, and amplifiers, all of which are fabricated in 1.5 μm standard CMOS technology.

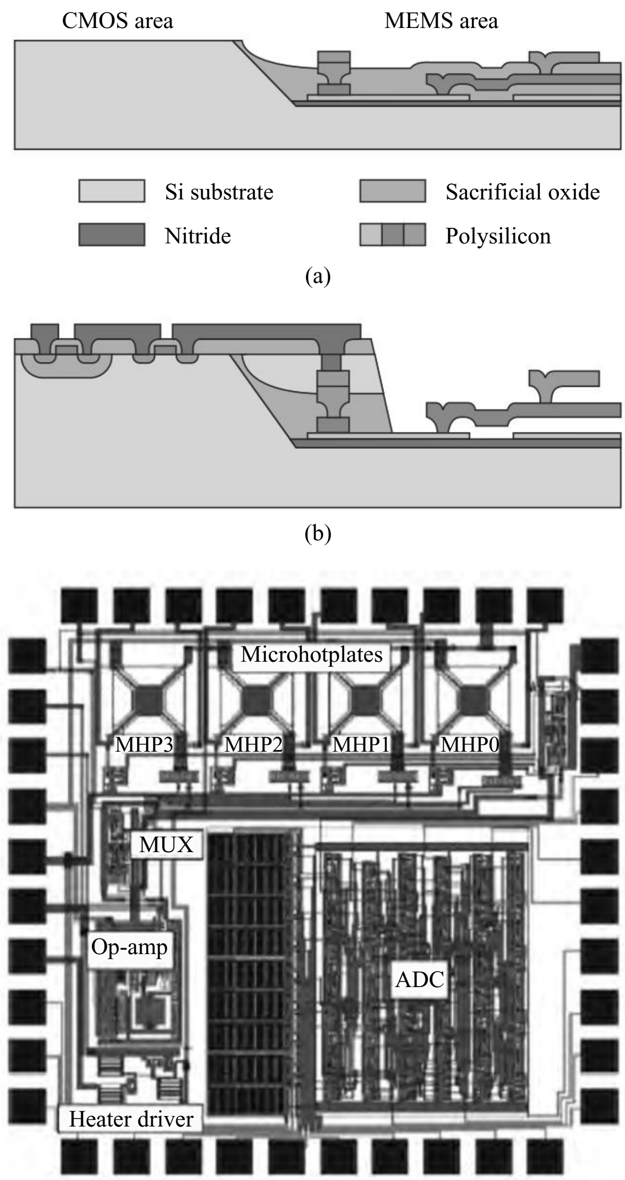


Fig. 4.1 *(a–b) An understanding of system-on-a-chip view (Source: J. H. Smith et al.), (c) Layout of an SoC gas sensor (Source: Solid State Electronics, 48 (2004) pp 1777–1781, Copyright: Elsevier Science)*

4.3 PASSIVE ELECTRONIC SYSTEMS

This section describes the fundamentals of PES and circuitry. These circuits are useful for interfacing and signal conditioning. *Signal conditioning* is a process of preparing the available signal into a desirable one. In essence, signal conditioning transforms a raw signal into a refined or robust signal that lends

itself to be used as conveniently as desirable. Important signal conditioning functions are rectification, amplification, filtering, Analog to Digital (AD) conversion, Digital to Analog (DA) conversion, isolation and balancing. More details follow (Some important electronic circuits are provided in Appendix D).

4.3.1 Rectification

The unidirectional flow action is the basis of *rectification*, a process by which an alternating current (AC) is transformed or rectified into a direct current (DC). When only half of the input signal appears at the output the circuit is termed a half-wave rectifier. On the other hand, the full-wave rectification circuit rectifies the negative portion of the input signal along with the positive portion.

4.3.2 Clipping and Clamping

A *clipping* circuit transmits an arbitrary signal within certain limits. The signal above and below the limits are suppressed. For this reason clipping circuits are also referred to as limiters. In another way it can be said that clipping circuits limit the swing of a signal. *Clamping* circuit, on the other hand, shifts the input levels (voltage or current) without changing the shape of the signal. Depending on the requirement, the input signal is shifted either up or down, keeping the range constant. Clamping circuits are of two types, namely *positive clamper* or *negative clamper*.

4.3.3 Amplifier

A circuit that accepts an input signal and produces an output signal in the same way as the input but with a larger magnitude is known as an amplifier circuit. There are various types of amplifiers such as voltage, current and power amplifiers. In practice, the ratio of output to the input of the amplifier is called gain of that amplifier. In order to achieve the required gain, sometimes two or three stages of amplifier circuits are designed. This kind of amplifier is referred to as a *multistage* or *cascade* amplifier. In such designs the output of the first stage is fed to the input of the second stage and so on. The overall gain of such a multistage amplifier is approximately equal to the product of the individual gains $G = G_1 \times G_2 \times \ldots G_n$, where G is the overall gain, G_1, G_2, etc. are the individual gains (Fig. 4.2).

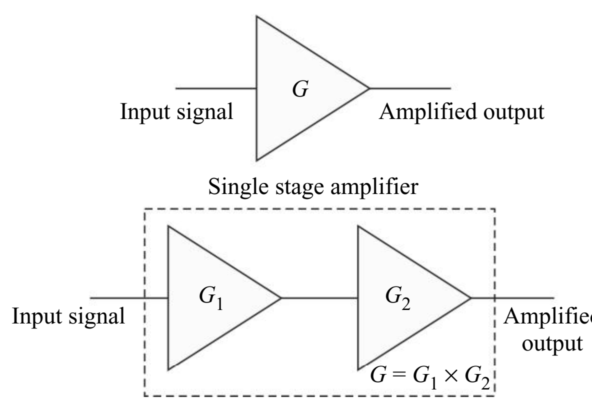


Fig. 4.2 *Schematic of an electronic amplifier*

4.3.4 Filter Circuits

Circuit that pass certain frequencies and attenuate (or eliminate) others are called filter. There are various types of filter circuits categorized by respective *frequency response curve*, which is a plot of the gain versus frequency, where the output and input may be either a voltage or a current. In order to compare the filter circuits, one type of input signal, usually a sinusoidal type, is taken. Besides low-pass and high-pass filters, there exist *band-pass*, *band-reject*, *narrow-band*, and *notch filters*. A band-pass filter passes one band of frequencies while rejecting both higher and lower frequency components. A filter that rejects one band of frequencies while passing both higher and lower frequencies is called a band-reject filter. Narrow-band filters have very narrow bandwidth. Notch filters pass almost every

frequency except for a few components. Band-reject filters and notch filters are synonymous, depending upon the frequency range with which their names appear.

Every filter has a BW. The bandwidth of a filter determines what frequency components it can allow to pass through without attenuation. In practice, the bandwidth is defined as the difference between the *upper cut-off* frequency, f_U and the *lower cut-off* frequency f_L, in the frequency response curve. The frequency response curve is plotted in dB scale. When plotted in dB scale it is called a *Bode plot*. The f_U and f_L are the frequencies at which the output power is 50% of the maximum possible output power. These are the two frequencies at which the gain is approximately 71% of the maximum gain. If the maximum gain is normalized to 0 (zero), then the gain at these points will be –3 dB. For this reason the lower cut-off and upper cut-off frequencies are called –3 dB points. The practical definition of the bandwidth of the filter becomes the frequency range (two extreme points in the frequency domain) of the filter within which the output power varies from 100% to 50%. When the filters are designed using capacitors and inductors, they are referred to as *passive filters*. A smoothing capacitor at the output of the rectifier circuit can be treated as a low-pass filter. Filters that are designed either using a transistor or OPAMP (Operational Amplifier) are called *active filters*.

4.3.5 Isolator

An isolator is a circuit that allows signal to be transferred between two circuits or systems, while keeping those circuits or systems electrically isolated from each other. The output impendance of the previous circuit or the input impendance of the next circuit may disturb the signal transmission between the two circuits. Isolation is important in order to avoid the effects of electrical loading between the two circuits. The gain of the isolator is unity. The output impendance of the isolator is very high; so it makes the following circuit electrically isolated from the previous one. The isolation may be achieved by utilizing optical signals. The electrical signal at the output of the previous circuit is converted to light signal then the converted light signal again converted to get back the electrical signal by using a photodevice satisfying the sanctity of isolation.

4.3.6 Instrumentation Amplifier

For high precision and accuracy applications differential amplifier configuration is used. The instrumentation amplifier (IA) is a differential configuration that uses three amplifiers as shown in Fig. 4.3. IA is usually used in the bridge circuit described below.

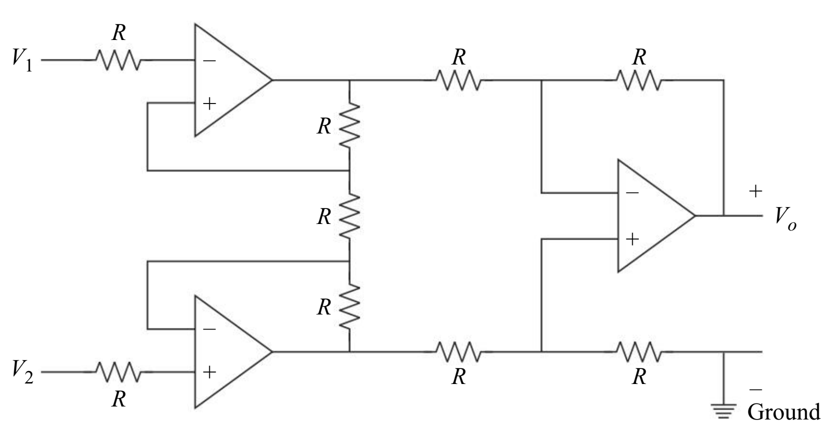


Fig. 4.3 *A typical instrumentation amplifier*

4.3.7 Bridge Circuit

A *bridge circuit* is an important signal conditioning circuit, which forms the basis of capturing a very small fraction of change, typically the resistance. The focus is given on *Wheatstone bridge*. Typically a change of resistance in the order of 0.0005 Ohm is detected. It is not practicable to measure the change of resistance of this order by any other means. In conjunction with IA the measurement is customarily carried out using a Wheatstone bridge.

Wheatstone bridge is a four-arm, four-terminal resistance-measuring electronic network (Fig. 4.4). The two junctions *A* and *B*, are for input voltage and the other two junctions *C* and *D*, are for output measurement.

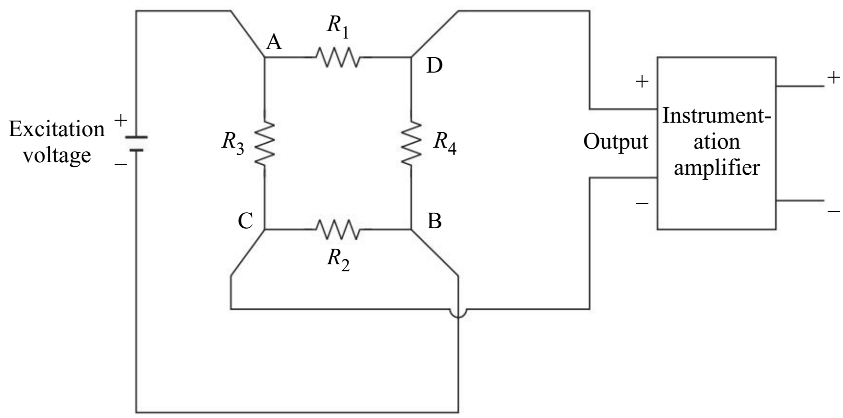


Fig. 4.4 *A Wheatstone bridge based signal conditioning circuit*

A DC voltage source called excitation voltage is applied between two input terminals. The output voltage is measured across the other two terminals (i.e. C and D). When the output voltage is zero, the bridge is said to be in a *balanced* condition. One of the arms called the *active arm* constitutes a resistive transducer, whose value change according to the physical measurand. The other arms of the bridge are simply resistors with fixed resistance equal to that of the active arm. When the resistance of the active arm changes in accordance with the measurand (say, strain) the previously balanced bridge will be unbalanced. The amount of unbalance is a measure of the measurand. The unbalance effect will appear as a voltage across the output terminals. The voltage is proportional to the change in resistance. The voltage is then amplified by the IA and can be calibrated to reflect the measurand. The expression for the output voltage V_{out} is given by,

$$V_{out} = \frac{1}{4}\frac{\Delta R}{R}V_{in} \tag{4.1}$$

where, V_{in} is the excitation voltage and R is the resistance of the arms and ΔR is the change in resistance. If the resistors have power dissipation characteristics, instability in the circuit will arise. Because of its high sensitivity, Wheatstone Bridge is also very advantageous for the measurement of inductance and capacitance.

4.3.8 Comparator

Comparators are widely employed in applications where some sort of signal comparison is required. It is an electronic circuit that typically has a threshold or reference setting to which the voltage level of a

signal at hand is compared. If the signal to be compared is greater than the reference setting then the comparator circuit provides an output in terms of a pulse or step signal. Zero crossing detector (ZCD) is a special type of comparator that provides a pulse or step when the input signal passes the time axis (zero level).

4.3.9 Oscillator

Signal conditioning circuits include *waveform generators*. Generator that produce oscillatory signal with high degree of stability are called oscillator. Various types of oscillatory waveforms are required. The types of oscillatory waves are sinusoidal, triangular, impulsive, square-wave, saw-tooth and staircase. They generate a changing output that repeats regularly at constant intervals. In communication systems, oscillatory signals are required for modulation and demodulation of original *baseband* signal. There are other applications as well. All oscillator circuits are positive feedback systems in which the loop gain is unity. In order to get stable and accurate oscillatory signal output, a piezoelectric quartz crystal is often used. The property of the quartz crystal is that when a changing mechanical stress is applied to the crystal, an oscillatory voltage is developed, whose frequency is the same as that of the mechanical stress vibrations. The highest vibration occurs at the natural frequency of the crystal. The frequency depends on dimensions, type of cut, thickness and temperature.

An oscillator whose output frequency can be changed by changing its input voltage is called voltage-controlled oscillator (VCO). VCO is used as the feedback circuit for the Phase Locked Loop (PLL) system in which a phase comparator exists. That is a phase comparator along with VCO, constitute a PLL. When properly set, the PLL can work as a demodulator for the frequency modulated signal. For more details on PLL and the demodulation technique refer Principle of Communication Systems, Taub and Schilling.

4.3.10 Sample and Hold Circuits

Sample and hold circuits (S/H) are very important in the context of analog to digital (AD) conversion of the signal. An S/H circuit is one that samples the analog signal at a particular instant and retains the value for a specified time for subsequent process such as coding.

4.3.11 Analog to Digital Converter (ADC)

Digital techniques and methods provide many added benefits over its analog counterpart. In order to deal such signal digitally there is a requirement to convert the available analog signal to a digital equivalent. The process is called Analog to Digital (AD) conversion. The AD conversion is achieved through three sub processes such as sampling, quantization and coding. Sampling and quantization can be achieved by S/H circuit.

There are various ways of realizing an AD conversion circuit, called AD Converter (ADC). Basically, two types of realizable circuits are encountered: counter-based technique and successive approximation technique, respectively. Figure 4.5 shows a schematic circuit diagram of a counter based ADC. There are three blocks: (i) A *comparator* circuit that gives an output, i.e. a pulse when the inverting terminal exceeds the input analog signal applied to the noninverting terminal, (ii) A Digital to Analog Converter (DAC) that converts digital signal to analog signal, and (iii) A counter that counts from zero to upward. The output of the comparator is connected to the input of the counter and the output of the counter is connected to the DAC. The digital output is taken from the counter only. With this description let us understand how the circuit works.

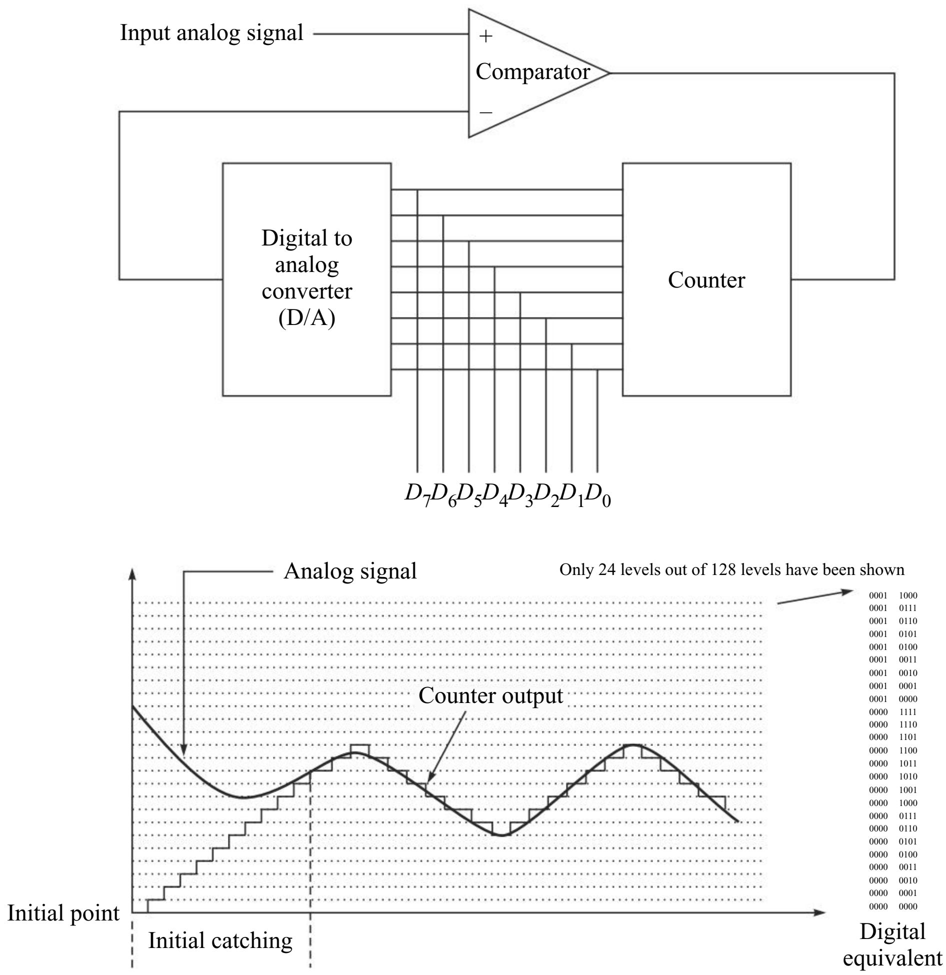


Fig. 4.5 *Counter based analog to digital (AD) conversion*

Initially, counter and the output of the DAC was reset to zero. When analog signal is applied to the noninverting terminal of the comparator, it produces a pulse, which is fed to the counter. The counter will be incremented by one from zero. The counter output, which is digital signal (a single pulse at the moment), is now converted to analog signal for comparison. If the input analog signal is still higher, another pulse will be generated at the comparator output, which in turn increment the counter a step. The process continues until the analog input is just smaller than the DAC output. Then the output of the counter becomes an equivalent of the input analog signal. After the initialization the equivalent digital output of the analog signal is acquired only when the output of the DA is greater than the input signal, and it may take some time, which depends on the input level and the number of bits within a counter. The counter based technique therefore inherits *initial catching problem.*

Successive approximation technique overcomes the initial catching problem, encountered in the counter based approach. Figure 4.6 shows the circuit diagram of an AD conversion based on successive approximation technique. It is composed of four functional units. The comparator, that

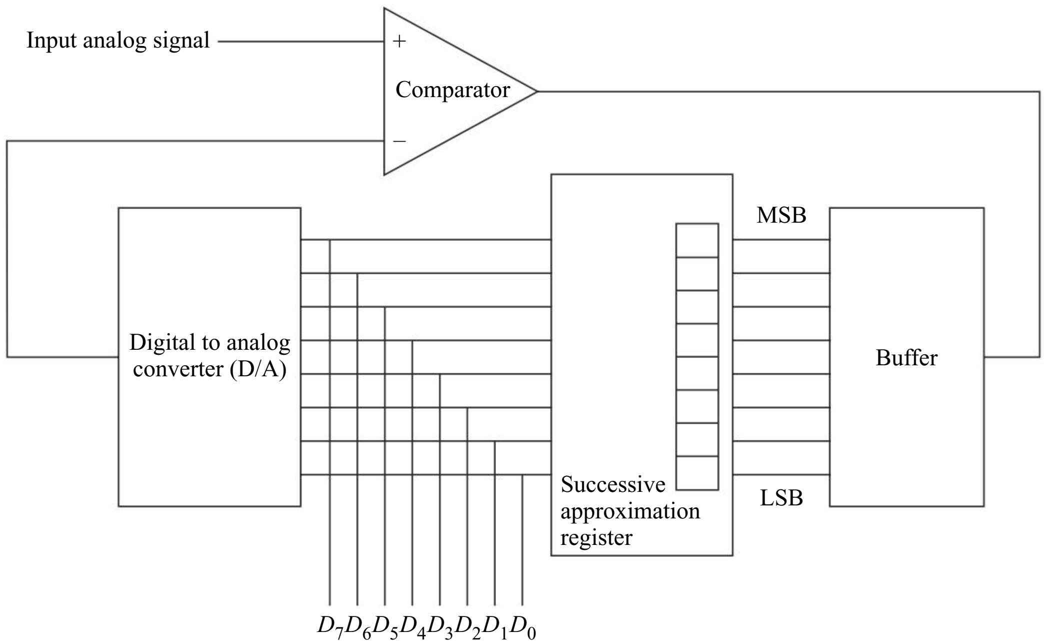


Fig. 4.6 *Successive approximation technique*

produces a pulse at the output, if the input signal (non inverting terminal) is greater than the signal at the inverting terminal. The output is fed to a register, called buffer. The buffer is connected to another register called Successive Approximation Register (SAR). The output of the SAR is the digital equivalent of the analog input. The last block, as before is a DAC, whose input is from SAR and the output is connected to the inverting terminal of the comparator. The circuit works as follows.

Initially the circuit is in reset state. This implies that the output of the comparator, content of the buffer and SAR and the output of the DAC, all are zero. When the analog signal is applied at the noninverting terminal, the comparator provides a pulse. In response to this the Most Significant Bit (MSB) of the SAR is set to '1' (In counter based method the Least Significant Bit was set) and the output of the SAR becomes a digital signal of value '1000 0000' (If it is an 8-bit ADC). The digital value is then converted to analog signal by DA block for comparison. If the analog equivalent of '1000 0000' is greater than the input analog signal, then the MSB of the SAR will remain set, else it will be reset to '0' and the next MSB will be set to '1', i.e. the new SAR output will be '0100 0000'. The analog equivalent of this new output will again be compared with the input analog signal, and if the analog equivalent of '0100 0000' is greater than the input analog signal, then this bit of the SAR will remain set, else it will be reset to '0'. The setting or resetting process based on comparison continues for all the bits. Once all the bits of the SAR are covered successively, the conversion is said to be completed.

A successive approximation based ADC takes as many clock cycles, as there are output bits to accomplish a conversion. In real implementation an ADC consists of a comparator, a DAC and a logic controller (LC). The LC uses the DAC to produce a voltage, and then and there observe the output of the comparator to check whether or not it is over or under the unknown input voltage at the moment. The LC then makes an adjustment, and observes again. It always takes exactly the number of steps, which is equal to the bit number. As an example, an 8 bit successive approximation ADC takes 8 steps to

perform the conversion whereas a 16 bit ADC takes 16 steps, and so on. The following technical parameters are considered while knowing more about the ADC.

- *Resolution:* The resolution is defined as the number of discrete values the ADC can produce and is expressed in bits. As an example, if the number of discrete levels in a conversion proces is 1024 then the resolution is 10 bits. Most converters sample with 6 to 24 bits of resolution, and produce fewer than 1 megasample per second.
- *Accuracy:* Accuracy also depends on the number of discrete levels and hence on the quantization error. Thus accuracy depends on the error in the conversion. The error is measured with reference to step size of the least significant bit. In the above example of a 10-bit ADC, an error of one LSB is 1/1024.
- *Sampling rate and aliasing:* The analog signal is continuous in nature. In some applications the converted digital signal is again converted to its analog equivalent after being processed digitally. The entire processs is done in such a way that the reproduced analog signal should be as close as to the original analog signal as possible. To achieve this there is a requirement to sample the original signal at a rate which is higher than twice the highest frequency component present in the original signal, otherwise aliasing will result. Therefore, before using an ADC the frequency content of the signal at hand has to be known. If the ADC is interfaced to a sensor device used to provide the feedback signal from a controllable plant then the *control bandwidth* of the plant must be known beforehand.

4.3.12 Digital to Analog Converter (DAC)

Mainly two techniques are employed as far as DA conversion is concerned. The circuits shown in Fig. 4.7 and Fig. 4.8 are called *binary resister-based* and *R-2R ladder-based* DAC, respectively. Both types commonly use an adder circuit (summing amplifier) with different implementation techniques. The latter one is better in terms of reliability, which can be understood once the operations of both the circuits are known. In Fig. 4.7, the voltage V_d is called the imposed signal, which is kept constant. The switches S_0 through S_7 represent electronic switches, which can be realized using transistors. In order to avoid the complexity, the detailed electronic circuit with regard to switches, has not been shown. However, bear in mind that the switches are operated (close and open) through 8-bit (a byte) input digital data, which needs conversion. Switch, S_0 is operated by the data bit D_0, switch S_1 is operated by data bit D_1 and so on. Logic '0' and logic '1' cause the switch to open and close, respectively. As an example, a digital data "01011101" at the input of the summing amplifier makes, S_0, S_2, S_3, S_4 and S_6 to close and S_1, S_5 and S_7 to open, respectively. Note the values of the resistances, which are in the binary order. When a switch is closed the current passes through the respective resistors. The value of the current depends on the value of the resistance. Since the values of the resistances are in binary order, so are the currents. When all switches are open, no current passes through and the output is zero. That is $I_i = 0$. When all switches are closed the current through the respective resistors are,

$$
\begin{aligned}
&I_0 = \frac{V_d}{128R}; \quad I_1 = \frac{V_d}{64R}; \quad I_2 = \frac{V_d}{32R}; \quad I_3 = \frac{V_d}{16R} \\
&I_4 = \frac{V_d}{8R}; \quad I_5 = \frac{V_d}{4R}; \quad I_6 = \frac{V_d}{2R}; \quad I_7 = \frac{V_d}{R}
\end{aligned}
\tag{4.2}
$$

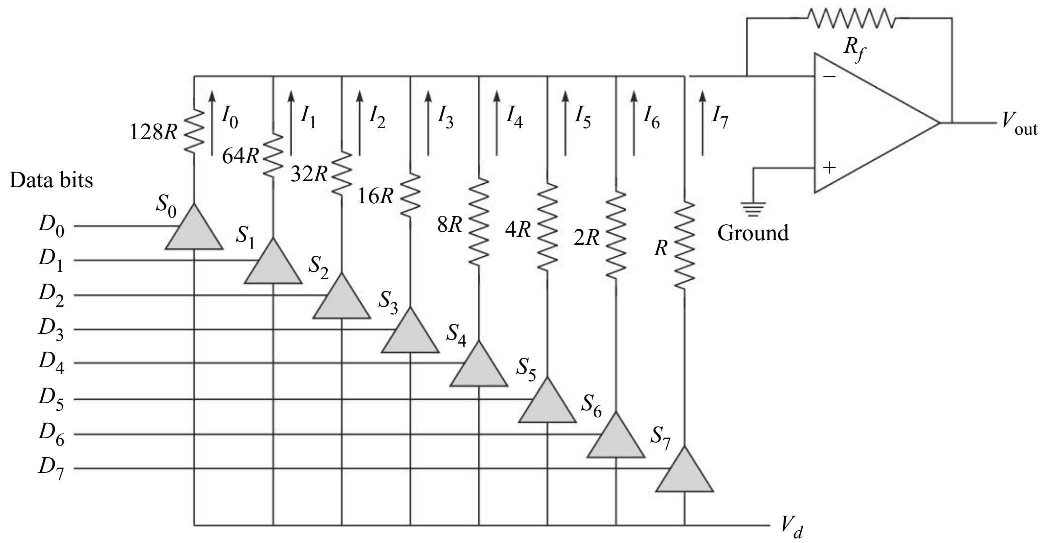


Fig. 4.7 *Binary resistor based Digital to Analog Converter (DAC)*

where I_i is the current through the *ith* resistor. If the value of imposed voltage is 5 volts (most preferable value), then the binary weighted currents due to the most significant bit and that due to least significant bit become,

$$\begin{aligned} I_7 &= 5/R \\ I_0 &= 5/128R = 0.0390625/R \end{aligned} \tag{4.3}$$

Thus, the current through any resistor represents the weighted value of the bit position. For unique combinations of input bits the summing circuits will produce a unique analog output, since the circuit sums up input currents. The number of combinations in this case is 2^8 (i.e. 128).

Figure 4.8 uses another technique called R-2R ladder method of DA conversion. In this case, the imposed signal is a current signal. The switch S_0 through S_7 are operated by the data bits. Switch, S_0 is operated by the data bit D_0, switch S_1 is operated by D_1 bit of the digital data and so on. A "0" causes the switch to open (ground) and a logic "1" causes the switch to close. As an example, a digital data "10100010' at the input of the summing amplifier makes, S_1, S_5 and S_7 to close and S_0, S_2, S_3, S_4 and S_6 to open, respectively.

Notice the value of resistances and their connections. Vertical resistors have value twice that of the horizontal resistors. Because of R-2R configuration the current is divided into two halves at each node, *N*. The current through any vertical resistor represents the weighted value of the bit position. For unique combinations of input bits the summing circuits will produce a unique analog output, since the circuit sums up input currents. For simplicity, assume that the value of the imposed current source is 256 mA. Now,

$$I_0 = \frac{I_d}{256}; \qquad I_1 = \frac{I_d}{128}; \qquad I_2 = \frac{I_d}{64}; \qquad I_3 = \frac{I_d}{32}$$

$$I_4 = \frac{I_d}{16}; \qquad I_5 = \frac{I_d}{8}; \qquad I_6 = \frac{I_d}{4}; \qquad I_7 = \frac{I_d}{2}$$

The R-2R ladder based DAC is reliable because there are only two values of resistances, i.e. R and 2R, required for implementation. However, in former case (binary-resistor) eight resistance values of

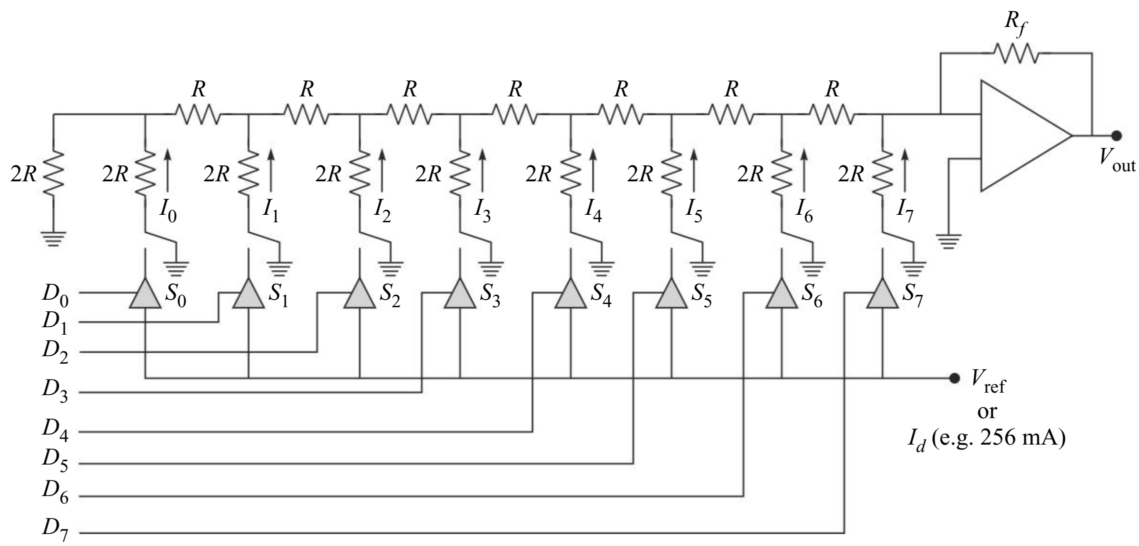


Fig. 4.8 *R-2R ladder based Digital to Analog Converter (DAC)*

binary order are required. In practice, such precise values of resistance are difficult to obtain as compared to only two values as in case of R-2R implementation.

4.3.13 Switches

In networking system a device that filters and forwards data packets between the Local Operating Network (LAN) segments are conventionally called switches. In these applications switches provide services of the ISO/OSI (International Standards Association/Open System Interconnection) seven-layer model. The principle of operation is that the input data are to be forwarded to the appropriate destination. This is a kind of channeling scheme based on some form of switching operation. That is, the input data need to be switched to a particular channel that is connecting the destination. Although, channels can be co-axial cable, waveguide, optical fiber, or any media in which the data can be transmitted but the switching operation is carried out by the use of *switches*. Switches are realized by the use of electronic devices such as transistors and diodes. The most essential requirement of switches is that the switching action should be faster. Mostly, FET (Field Effect Transistors) and PIN (positive Intrinsic Negative) diodes are used in order to realize a switch due to the reason that these devices have relatively good switching capability. Furthermore, they are reliable. In the sequel, principles of operation of switches are presented.

Figure 4.9 shows three forms of switches. Figure 4.9(a) is a simple type. The architecture has one common terminal called COM terminal and N switched terminals. The appropriate switch is made close to channelise the signal through any output terminal. Figure 4.9(b) shows a multiplexer switch that uses 4×1 building blocks. Note that only one signal can be routed at any given moment.

Figure 4.9(c) shows another technique for switching operation. The RF_{in} signal can be routed to any out of six RF_{out} terminal by using control signal. The switches (FET or PIN diodes) are controlled by the control signal through in-housed decoder circuits. Input buffer (a temporary memory cell) is needed to store the input control signal. The switch closes and opens according to the control signals.

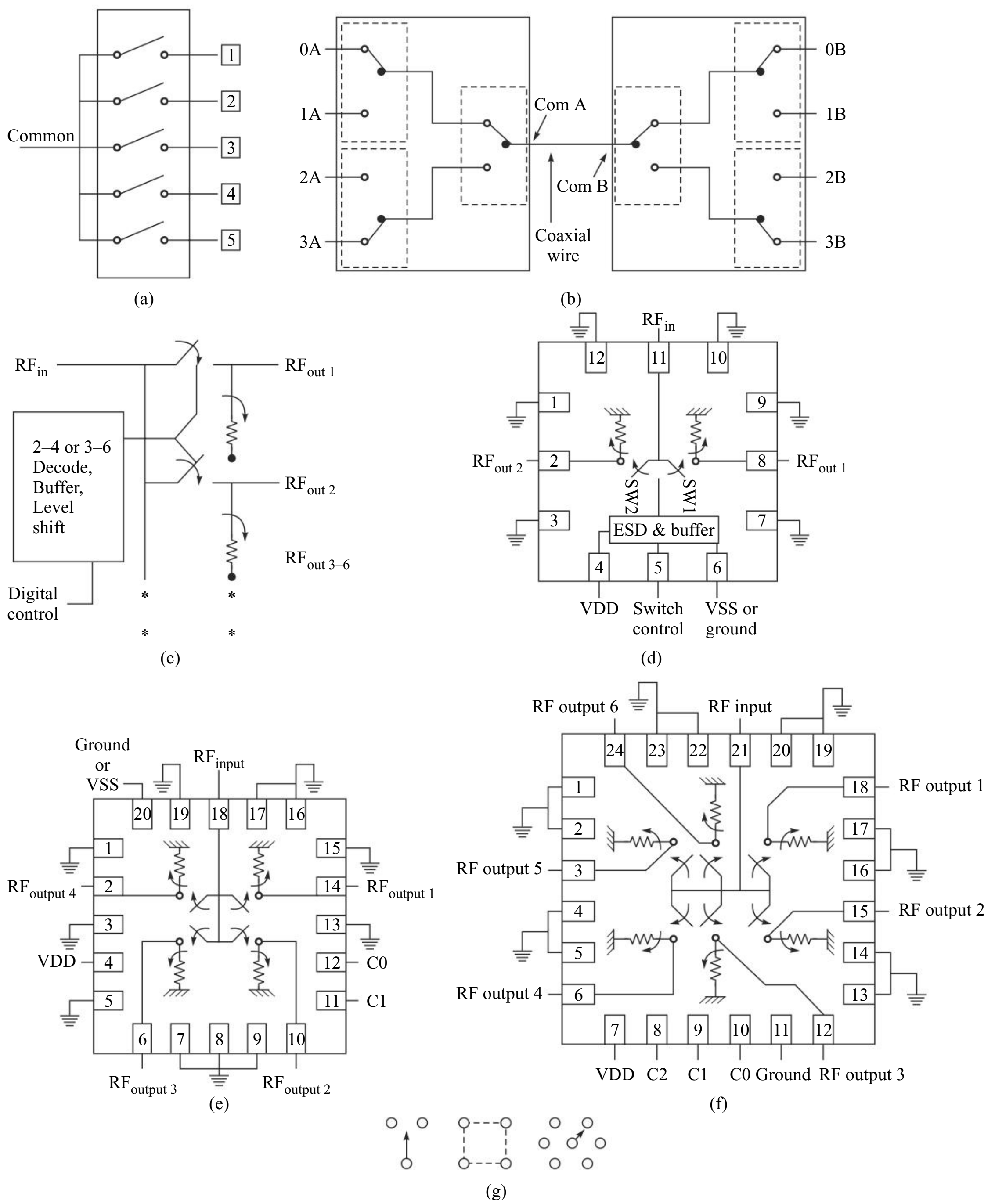


Fig. 4.9 *(a) Schematic diagram of a typical simple switch; (b) Cascaded matrix switch; (c) Techniques used in the IC based RF switch operation; (d) Single pole double throw switch; (e) Single pole four throw switch; (f) Single pole eight throw switch (g) Switching behavior*
Courtesy: Honeywell Inc.

The control signal that makes the switching action possible is also called actuation signal. The switch itself is sometimes called actuator. One way of classifying the switches are as follows.

- Single pole single throw (SPST)
- Single pole double throw (SPDT)
- Double pole single throw (DPST)
- Double pole double throw (DPDT)
- Single pole four throw (SP4T)
- Single pole six throw (SP6T)
- Single pole twelve throw (SP12T), etc.

Figures 4.9(d), (e) and (f) show schematic diagram of single pole double throw switch, single pole four throw switch and single pole eight throw switch, respectively. Figure 4.9(g) illustrates three types of switching behavior such as throw, transfer and multiposition. Important performance specifications to consider for switches are,

- Frequency range
- Switching speed
- Insertion loss
- Isolation
- Impedance
- VSWR (Voltage Standing Wave Ratio)

The frequency range is the desired signal frequency range that the switch can handle. The switching speed is the speed at which the switch can operate. Faster switching is always desirable. Insertion loss is defined as the total transmission loss resulting from the insertion of a device in the transmission system. The insertion loss is the power loss and defined as the ratio of output signal power to the input signal power of the switching device. If the input terminal is defined as port A and the output is defined as port B, then they must be isolated to each other. The isolation is a measure of degree of attenuation of the signal at port A measured at the B port with all other ports properly terminated. The isolation is always expressed in dB. Typical input and output impedance are either 50 ohms or 75 ohms, depending upon the applications. Voltage Standing Wave Ratio (VSWR) is a measure of signal reflection. Note that a switch is connecting two systems: the previous and the next system. The impedance of these two systems may not match. The signal may reflect if the impedance of the two systems mismatch. The range of VSWR is 1~ ∞ and is a unit less number. A value of one indicates that all the energy passes through, while any other value indicates that a portion of the energy is being reflected.

4.3.14 Phase Locked Loop (PLL)

The PLL is a useful electronic circuit which contains a phase detector and a VCO (Voltage Controlled Oscillator) (Fig. 4.10). As can be seen, the system is a feedback system. The phase detector unit compares two input frequencies, generating an equivalent output voltage that is a measure of their phase difference. Then the output is fed to the VCO, which generates frequency f_{VCO}, corresponding to the input voltage level. If input frequency $f_{IN,}$ is not equal to f_{VCO}, then the phase-error signal causes the VCO frequency to deviate in the direction of f_{IN}. If conditions are right, the VCO will quickly lock to the input frequency maintaining a fixed relationship with the input signal. The PLL can be used advantageously for demodulation of the baseband signal from the frequency modulated (FM) signal.

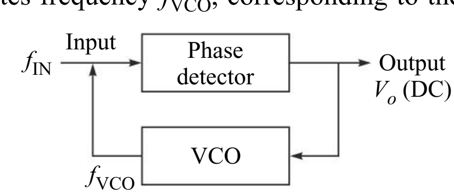


Fig. 4.10 *A PLL system*

4.4 PASSIVE MECHANICAL SYSTEMS (PMS)

MEMS PMS facilitate mechanical advantage. Mechanical advantage can be calculated by dividing the load by the effort. In general, the mechanical advantage of a PMS is the ratio of the output and input forces and apparently is a number that is greater than unity. For instance, gears make it possible for changing the rate of rotation. They can even change the direction of the axis of rotation and can convert rotary motion to linear one. The rest of this chapter describes fundamentals of mechanisms in relation to PMS and components.

4.4.1 Bearings

The bearing is a passive component that reduces frictional losses as surfaces side past one another. The function of a bearing is to reduce the friction between a fixed and a moving surface, and also to carry a load. We distinguish mainly five important types of bearings. They are, slide bearing, journal bearing, rolling element bearing, magnetic bearing and molecular bearing.

Slide Bearing A slide bearing, as the name implies, facilitates linear motion between a load and a support. These bearings are needed whenever one part slides against another. Figure 4.11 shows a typical illustration that uses slide bearing.

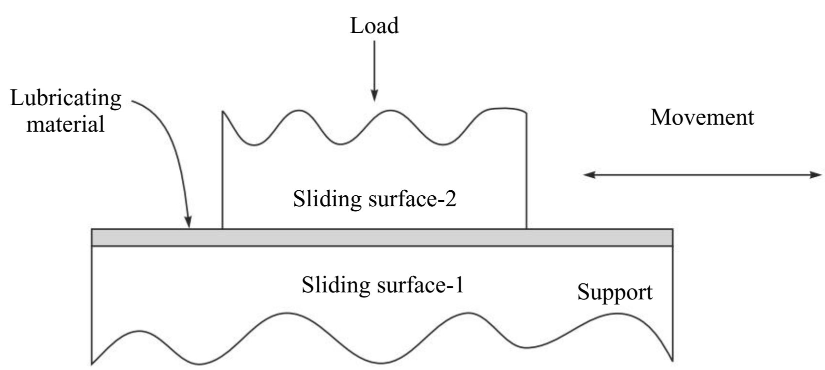


Fig. 4.11 *Slide bearing*

Journal Bearings The mechanism of journal bearings is similar to sliding bearing but in this case, a rotating shaft is involved. A journal bearing permits a load to exert radial pressure on a shaft.

Rolling Element Bearing Rolling element bearing is considered as one of the very important components and is essentially used for rotary applications. The surfaces of any object roll better than they slide because when objects slide, the friction between them causes a force that tends to slow them down, however, if the two surfaces roll over each other, the friction between them is greatly reduced. Note that rolling reduces friction related energy loss to a minimum, ideally to zero. The rolling element bearings consist of two circular metal rings and a set of rolling elements. One of the rings is larger than the other. The smaller of the two is called *inner ring* and the other one is referred to as the *outer ring*. The inner ring fits well within the perimeter of the outer ring. A fixed number of solid rollers are designed into geometric shapes and placed at equal intervals in the open space between the two rings. The power is transferred from one ring to the other through the roller elements with minimum loss.

There are many types of bearings as far as rolling elements are concerned. The geometric shapes of these rolling elements define the classification. Common rolling elements are balls, cylindrical, needle type and tapered rollers. Figure 4.12 shows the geometric shapes of different types of rolling elements.

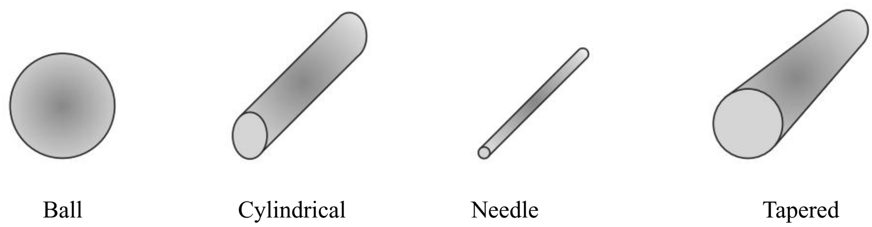
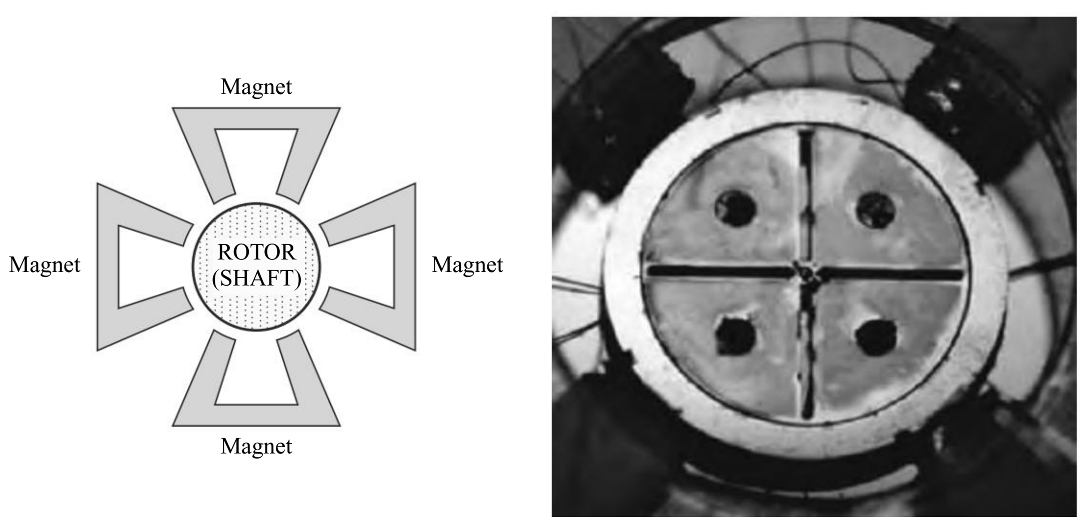


Fig. 4.12 *Typical roller elements used in the bearings*

Magnetic Bearings The limitation of friction losses in bearings is an engineering challenge. In magnetic bearings, mechanical friction losses are eliminated. Conceptual magnetic bearing is shown in the Fig. 4.13. It is composed of four numbers of horseshoe-shaped magnets. The four magnets may be arranged evenly around a rotating shaft. Each of the electromagnets can produce a force that attracts the rotor iron towards them. These bearings allow the rotating shaft to float on a magnetic field created by the electromagnet.

These bearings are used in lieu of journal and rolling element bearing and are essentially employed in high performance applications. The bearing is distinguished from other types by an almost complete absence of friction, since there is no mechanical contact. In addition, reliability can be increased because there is no mechanical wear. Besides the obvious benefits, magnetic bearings can allow some instability because any unbalanced magnetic field displaces the rotor position causing serious problems



(a) (b)

Fig. 4.13 *(a) Conceptual magnetic bearing composed of four discrete horseshoes, (b) SEM of an Assembled micro magnetic bearing*
Courtesy: Ghantasala, Copyright: Smart Mater. Struc. 9, 2000

to the system. The position of the rotor should be central to the geometrical axis of the system forever. Since the rotor shaft floats in the magnetic field, the effect of self-weight of the shaft must be considered. Magnetic bearings are generally open loop. In order to keep the position of the rotor fixed active feedback control is required for the bearings to operate stably. However, the requirement of feedback control sometimes brings inflexibility into the dynamic response of the bearings. MEMS magnetic bearing design is a challenge.

Molecular Bearing Because of the advent of *nanotechnology*, design of molecular bearings has already begun. It is worth pointing out that the word nanotechnology has become very popular and is used to describe the characteristic dimensions, which are less than about 1,000 nanometers. Nanotechnology is considered as the miniaturization to the extreme. Although nanotechnology is still at the rudimentary stage, potential research is underway all over the world. Atomic-scale machines are to be constructed to perform tasks in molecular environments. Currently, some researchers believe nanomachines may be only 10 years away. Nanotechnology will enable the construction of nanomachines (known as molecular machine), nanomanufacturing appliances, robot, computers, and much more.

The bearings are integral to most of the nano-world machines. Nanotechnology conformant molecular bearings are likely to be as ubiquitous in future molecular machines as conventional bearings are in today's macroscopic machines. Apparently, molecular bearing has become one of the most useful inventions in recent years. Bearing design takes significant work, because this needs several types of bearings for different applications. The simplest bearing is a single chemical bond, which is only useful for small loads. For larger bearings, two atomic flat sheets of some material, one larger than the other are bent to form cylinders and smaller are placed inside the larger. These types of bearings will not require lubrication.

4.4.2 Gears

Gear is another important component used to facilitate mechanism in terms of transformation of power and motion from one rotating shaft to another. They can help to increase or decrease force, torque or speed and can change the direction of the axis of rotation. They can also change rotary motion to linear motion. Figure 4.14 shows a simple but a basic gearing integration. The integration consists of two gears. One is larger and other is smaller in diameter. Each gear has been rigidly fitted with a shaft, to and from which the power is transmitted. Eventually, the two shafts couple each other through the gears. The gear has teeth around it. One of the characteristic feature of the gear is the number of teeth it has.

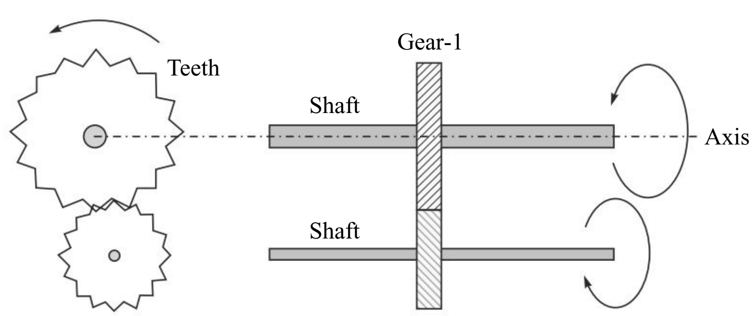
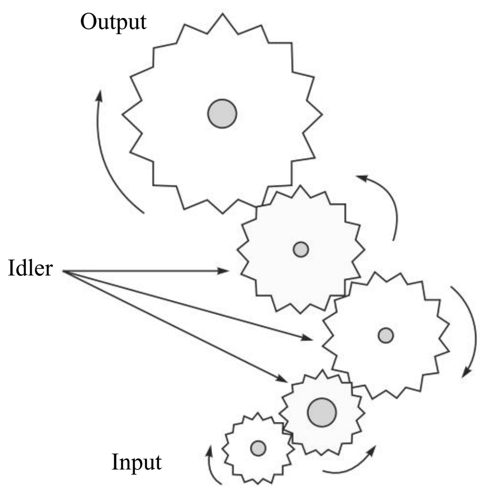


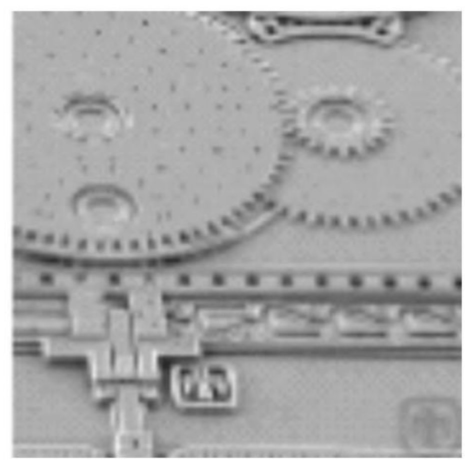
Fig. 4.14 *Basis of a typical gearing arrangement*

The gear that is turning the other one is called the *driver*, and the gear being turned is called the *follower*. Driving a small gear by turning a larger gear is called *gearing-up*, and driving a large gear with a small gear is called *gearing-down*. The ratio between the gears provides control of the speed and torque of the output shaft. On any gear, the ratio is determined by the distances from the center of the gear to the point of contact.

A single pair may be inadequate for certain sources and loads, in which case more complex combinations are necessary. The gears essentially work together at least in groups of two or more. More than two gears working together is called a *gear train*. They can mesh in many different ways depending upon the application requirements. One gear turns another, which may turn another, and so on (Fig. 4.15). The necessary gearing systems are designed to obtain the desired speed, direction, motion and torque. As before, the gear on the train to which the force/torque is first applied is called the driver. The final gear on the train to which the force/torque is delivered is called the driven gear or follower. Any gears between the driver and the driven gears are called the *idlers*.



(a)



(b)

Fig. 4.15 *(a) Schematic diagram of a gear train (b) SEM of an MEMS gear train*
Courtesy: Sandia Inc., USA

The speed and torque relationship through gearing mechanism can be understood as follows. A small electric motor, capable of rotating at high speed providing enough power, but not enough torque. If such a motor is to be used for electric screwdriver, then gearing mechanism must be incorporated into it, since driving the screw essentially requires high torque. The motor must have a very large gear reduction because it would need high value of torque in order to turn round the screws. In majority of applications, it is required to reduce the speed and to increase the torque. That is, the speed reducers are much more common than speed-up gear systems. The speed-torque relationships are such that sometimes the *speed reducers* are referred to as *torque amplifiers*. With a gear reduction, the output speed can be reduced while the torque is increased.

The way the teeth are designed classifies the gears. Basically, four types of gears, namely spur, bevel, helical and worm, are found in engineering applications. Spur gears are common in MEMS. A spur gear is one of the most important ways of transmitting a motion between two shafts lying parallel to each

other. The design is such that the teeth in the respective gears are always definite and regular. Clearance is the distance from the tip of a tooth to the circle passing through the bottom of the tooth space and measuring radial. In terms of load bearing capability, the correct clearance is vital to the motion of gears. These gears generally cannot be used when a direction change between the two shafts is required.

One of the basic relationships for a gear is the number of teeth, the diameter, and the rotary velocity. If one gear has 120 teeth and another has 20, the gear ratio is 6.1. A gear ratio is simply the relationship between how often the drive shaft turns to complete one full turn of the other or vice versa. When the smaller gear makes a complete rotation the larger gear only makes a lesser turn. The gear ratio is related to the ratio of speeds.

$$R = \frac{w_a}{w_b} \tag{4.4}$$

where, R is the gear ratio; w_a and w_b are the speeds of the driving gear and driven gears respectively. The velocity ratio of the driver and follower does not change by putting any number of gears between them. Therefore, the gear ratio is the ratio of speeds of the driver and driven gear. The gear ratio for Fig. 4.16(a) is,

$$R = \frac{w_1}{w_2} \times \frac{w_2}{w_3} = \frac{w_1}{w_3} \tag{4.5}$$

Figure 4.16(b) shows compound gears in which the two gears are on the middle shaft. Gears-2 and Gear-3 rotate at the same speed, since they are fixed to the same shaft. The gear ratio in this case is,

$$R = \frac{w_1}{w_2} \times \frac{w_2}{w_3} \times \frac{w_3}{w_4} = \frac{w_1}{w_4} \tag{4.6}$$

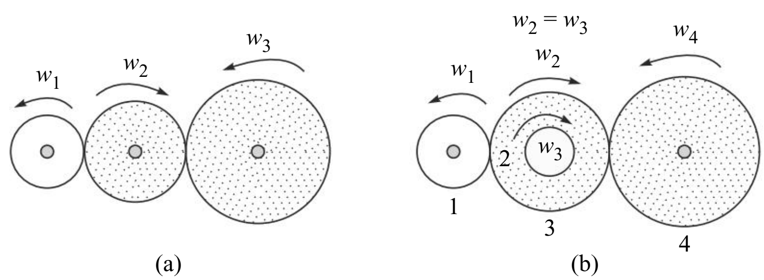
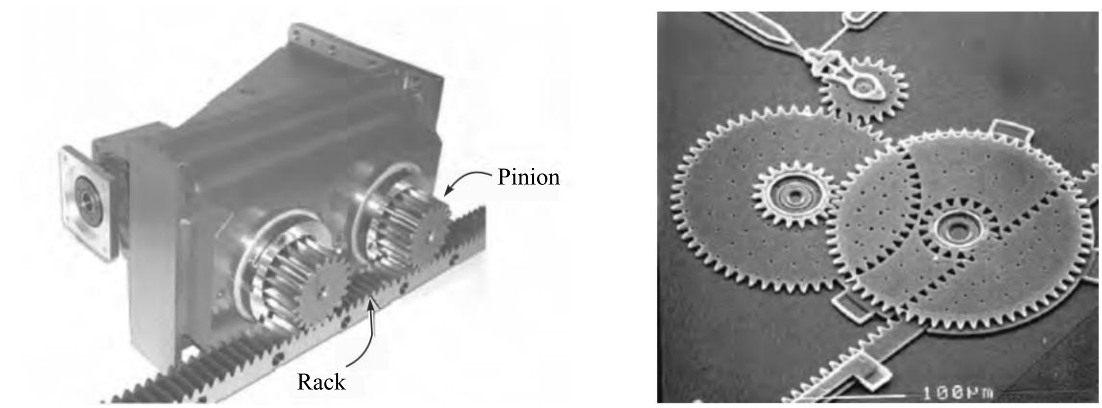


Fig. 4.16 *Illustration for gear ratio calculation*

4.4.3 Rack and Pinion

Rack and pinion driving system (Fig. 4.17) refers to a special type of gearing mechanism in which the rotational motion is converted to linear one and vice versa. As the name implies, rack and pinion consists of two major components; a *rack* and a *pinion*. The rack is a long piece of MEMS material that is flat. The flat side contains teeth running the length of the rack. The teeth are cut perpendicular to the edges of the rack. The other component, the pinion is like a gear that also has teeth on it. The pinion resides in a fixed position and drives the rack, but the reverse is also possible. Whatever may be the requirement, in summary, the rotary motion of the pinion is changed to transverse motion by the rack or transverse motion of the rack is changed to rotary motion by the pinion.

The teeth of the rack and pinion can be helical or straight. Compared to straight teeth, helical teeth can bear more loads, because the area of contact between the teeth of the rack and pinion is more. The



(a) (b)

Fig. 4.17 *(a) Traditional macro-level rack and pinion (Courtesy: Andantex Inc. USA) (b) An electrostatic micro engine output gear coupled to a double-level gear train that drives a rack and pinion slider. This gear train provides a speed reduction/torque-multiplication ratio of 9.6 to 1*
Courtesy: Sandia National Laboratory (Microelectronics Development Laboratory); *Source*—Smith, Sensors Expo, 1996

rack and pinions are made up of metal or hard plastic, depending upon the speed, load, torque requirement. For high precision positioning, where *backlash* cannot be tolerated, rack and pinion drive systems are designed with dual pinions. Typical applications include high-speed axis drives, positioning systems, traveling gantries and machining.

4.4.4 Ratchet, Pawl and Crank

The *ratchet* is an asymmetric mechanical component that allows something to turn in one direction only and it happens in a very jerky manner. In conjunction with another component called *pawl*, fundamentally, as if the ratchet works like a locking system. In the *reset* position, the pawl prevents the ratchet from moving backward. The ratchet is considered as another form of gearing mechanism. It is designed with suitably shaped teeth, receiving an intermittent circular motion from another member, called *crank*. Unlike gears, which can be used to speed up or slow down movement, the ratchet can only be used to slow the motion down.

A simple form of ratchet mechanism is shown in Fig. 4.18(a). For every 16 turns of the crank the ratchet wheel turns one complete revolution. Another type of ratchet-pawl mechanism can be seen from Fig. 4.18(b). In both the cases, the crank is oscillating in nature. As noticed, there are two pawls: one is flexibly attached with the crank (called primary pawl), and the other one is separately fitted (called supplementary pawl). The *supplementary* pawl prevents backward motion of the wheel. In Fig. 4.18(b) as the crank moves backward (counterclockwise), the attached pawl slides over the tooth. Then the crank is allowed to move forward (clockwise). This causes the wheel to move forward direction. The wheel is moved through a fractional part of a complete revolution, which only depends upon the motion of the crank. The supplementary pawl is there to prevent the backward movement of the wheel. Here ends the first cycle. When the crank moves back, the second cycle starts and the attached pawl again slides over the teeth while the wheel remains at rest because of the supplementary pawl. The crank is then ready to push the wheel on its forward motion as before. The amount of cyclic motion varies with

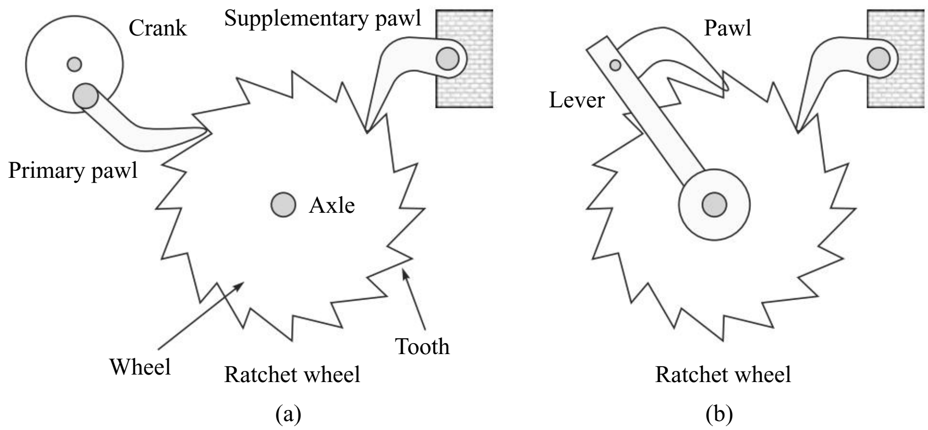


Fig. 4.18 *Ratchet, pawl and crank*

the pitch of the teeth. Designing smaller teeth could reduce this motion. Figure 4.19 shows a MEMS ratchet and pawl.

Fig. 4.19 *MEMS complex ratchet mechanism*
Courtesy: Sandia National Laboratory USA

4.4.5 Slider and Crank

The *slider–crank* is a linkage (Fig. 4.20) that transforms linear motion to circular motion or vice versa. The crank is a lever attached to a rotating shaft. It is really a *link*, which revolves relative to a frame. Various types of motion transformation can be achieved through this mechanism by designing the link properly. The link can convert a rotating motion into a reciprocating motion or vice versa. The process of determining

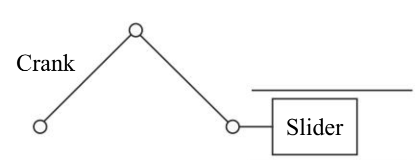


Fig. 4.20 *Slider–crank mechanism*

the exact relation between the reciprocating motion of the piston and the angular motion of the shaft for a given linkage mechanism is called slider–crank kinematic analysis. The slider–crank can also transform a rotation into a combined motion up and down and from side to side.

Often slider–crank linkage is called constant-force mechanisms when it operates in reverse manner. It produces a constant output force for a range of input displacements. Such mechanisms are important in applications with a varying displacement but a constant resultant force required. The constant force mechanism is displacement driven. The input is a displacement at the slider and the output is a force. Unlike a linear spring where the force increases as the displacement increases, the reaction force remains constant for various displacements. Mostly, these mechanisms have specific geometry and mechanical properties such as stiffness that cause the combination of energy storage and mechanical advantage to produce the necessary constant force.

4.4.6 Geneva Wheel

Figure 4.21 shows a *Geneva wheel*. Geneva wheel based mechanism is used for achieving intermittent motions. The component has two wheels, the upper and the lower wheel. There is a projection, called drive pin, mounted on the lower wheel. In this typical example, the upper wheel, called the Geneva wheel, has four slots, however, more slots can be designed depending upon the need. When the lower wheel rotates, the projection (drive pin) of the lower wheel is inserted into the slot, which makes the upper wheel to rotate in appropriate direction. The drive pin on the lower wheel engages itself with the slots on the Geneva wheel in such a manner that it is in position when the pin comes round again. In effect, the lower wheel drives the upper one. The rotational movement of the lower wheel is essentially continuous but the upper wheel only rotates step-wise, i.e. intermittently. The upper wheel makes one complete rotation corresponding to four complete rotations of the lower wheel, in this example.

4.4.7 Flexure, Anchor and Joint

In designing machine systems we are very much acquainted with joints. The joint is the point of connection between two mechanical elements. The elements could be a beam, a plate or post. Welding, nut and bolt, anchor and hinge make joints. Welding and nut-bolt joint methods are not applicable to

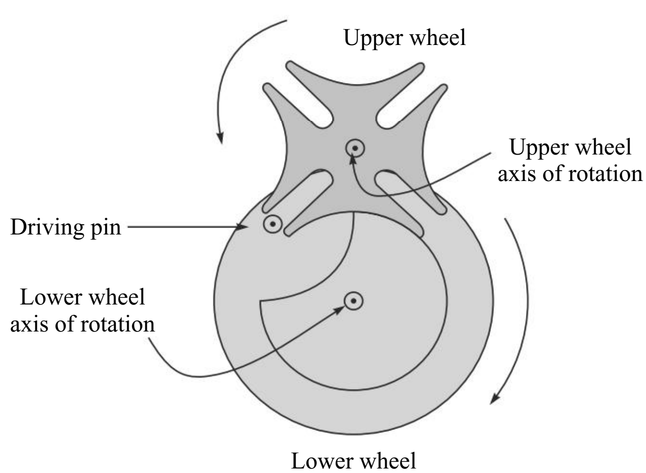


Fig. 4.21 *A typical Geneva wheel*
Source: Cabaret Mechanical Theatre *http://www.cabaret.co.uk/*

MEMS systems and devices. Mostly, hinges and anchors are used, but anchors are common. Basically, anchor is a mechanical connector that prevents something to move. More precisely, it is a segment at which part of an MEMS device is secured to the substrate to prevent part from moving. In many occasions, the immovable elements are not directly attached to the anchors, rather they are interfaced through spring or special cantilever arrangement. Such interfacing (anchor-spring/cantilever-element) is necessary when it is desired to bend one end of the element while keeping the other end (anchor end) fixed.

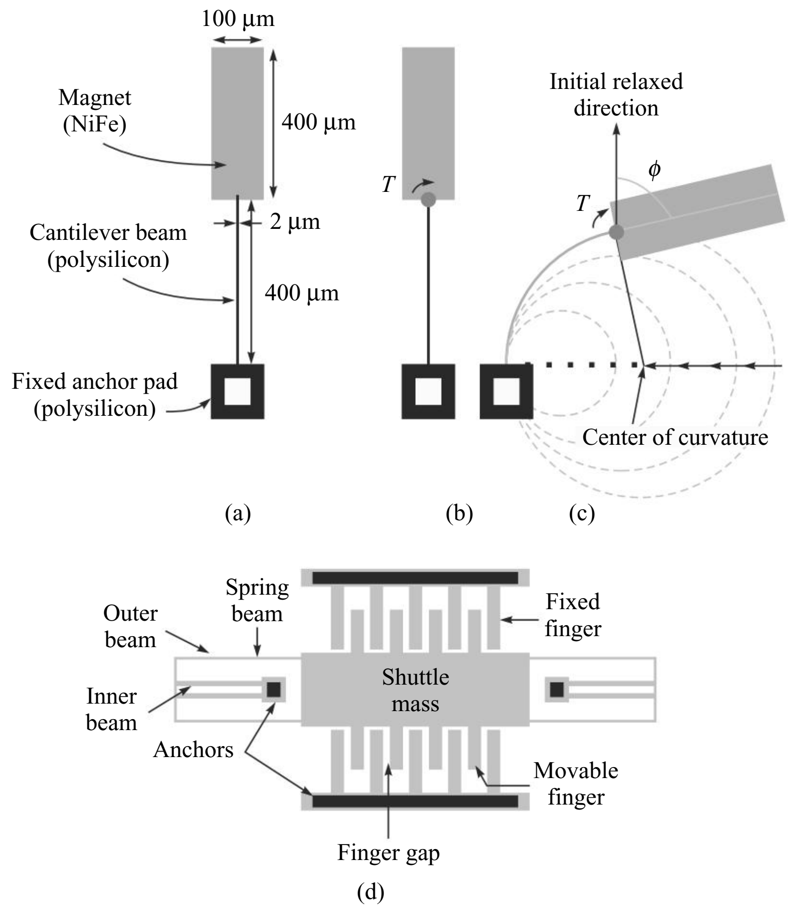
In another example, Fig. 4.22(d) shows a resonator, which is anchored at the appropriate place. Resonators have several applications including design of accelerometer. In fact, it is an important component of accelerators. Resonators (or cavities) are hollow bodies in which oscillating electromagnetic fields accelerate the particle beam. In radio and radar applications, a resonator behaves as a circuit that will resonate at a certain frequency or over a range of frequencies, when properly excited. The resonator was chosen as it possesses many of the basic structures that form the core primitives found in many other types of MEMS. The resonator structure is a mature device in the design of MEMS variants of which are used in commercial accelerometers, gyroscopes, etc. The top view of a surface-micromachined, electrostatic comb-drive actuated microresonator is shown in Fig. 4.22(d). Most of the resonator structure is suspended over the die substrate and is anchored only at the locations shown. The resonator makes it possible to transform shuttle movement to a capacitance change between the fixed and movable fingers. Shuttle movement can result from a potential difference between the shuttle and fixed fingers, or from an inertial force caused by an external physical parameters such as acceleration.

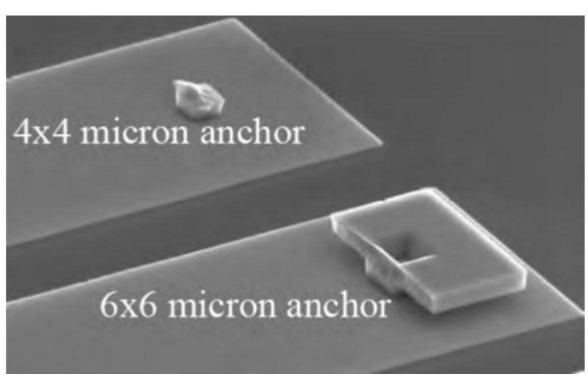
In a magnetic actuator shown in Fig. 4.22(a), one end of the magnet is anchored through a thin cantilever interface so that if a magnetic field is applied, the top magnetic element will be subjected to a torque allowing it to bend. The design can act as a magnetic actuator (more details in Chapter 8). The magnetic element along with the cantilever is known as a *flexure*.

Anchor-cantilever combinations are very useful to facilitate degrees-of-freedom (DOF). For example, in telecommunications micromirrors need to have 2-DOF in order to be actuated to extremely precise positions (Refer Figs. 7.16, 4.10 and 4.22(f)).

Traditionally, MEMS structures contain bendable joints, a severe limitation on mechanical design capabilities for many applications. At this point, it is therefore worthwhile to discuss the concept of *pin joint* (Fig. 4.22(g)) a preferred joint method mainly seen within MEMS design arena. Microstructures with rotatable joints, sliding and translating members, and mechanical-energy storage elements provide the basis for a more general micromechanical transducer-system design as such structures add important degrees of freedom to designers. A pin joint is composed of an axle around which a member (rotor) is free to rotate. Movement along the axle by the rotor is constrained by flanges. Pin joint is usually fixed to the substrate and fabricated using polycrystalline silicon. The rotor, axle, and flange are all made of polysilicon and deposited by LPCVD. Several different designs of pin joints have been developed using two layers of polysilicon. These pin joints will rotate as well as translate. Each type of design can be included in a number of micromechanism. Some pin joints allow for rotation while, at the same time, permitting translation across the silicon surface. These joints need to have a flange on the axle underneath the rotor to keep it in place. The axle can either be fixed to the substrate or else left free to translate across its surface.

Of primary interest is the design of various types of interfacing elements such as lever, cantilever or spring to encounter backlash, energy loss, stress and the way of achieving the movement (unidirectional or bidirectional). The design of flexure primarily looks at the bending moments in relation to geometry of

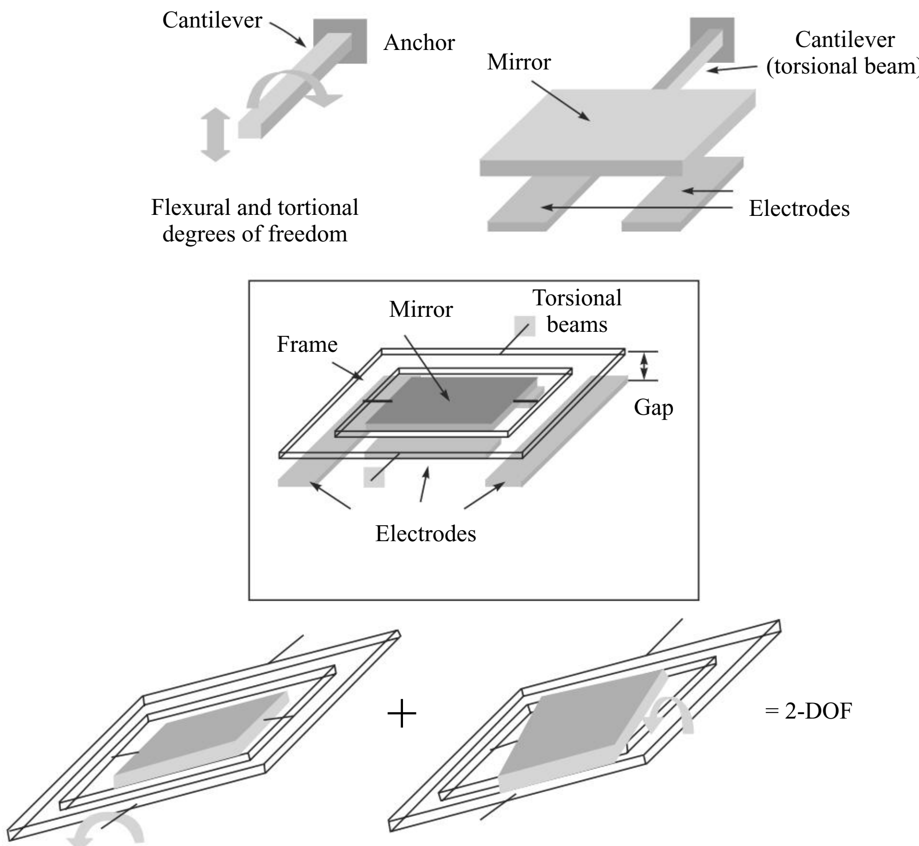




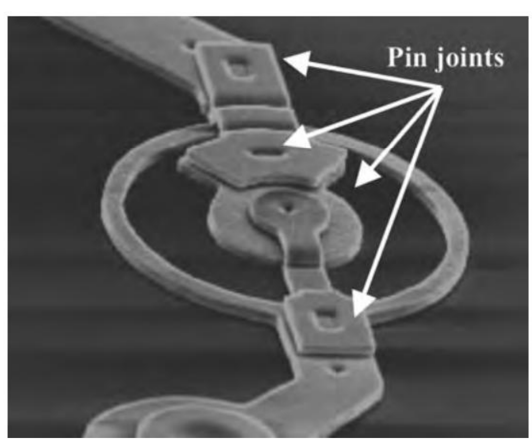
(e)

Fig. 4.22 *(a) A typical magnetic actuator that has a magnetic element (400 µm length, 100 µm width, 25 µm height), cantilever element (400 µm length, 2 µm width, 2 µm height) and a fixed anchor pad, (b) The magnetic actuator is subjected to a torque, T because of applied magnetic field, (c) The resulting bend, ϕ. (Courtesy:* Judy et al. U. of California, Berkeley, Solid-State Sensor and Actuator Workshop,

(Contd)



(f)



(g)

1994), *(d) electrostatic comb-drive actuated microresonator, (e) An SEM of an anchor, (f) Use of anchor for facilitating degrees of freedoms (DOF) (Courtesy:* UC Irvine Microsystems Lab.), *(g) A typical pin joint .*

the neck. Practically, the flexure utilizes thin necks. There are several ways to anchor the element through the neck. The design largely depends on two parameters; the geometry of the element to be anchored and the required bending range. The geometry of the element could be a plate or beam structure. When large bending or motion is necessary, spring type interfacing (neck) is inevitable. Figure 4.23 illustrates some commonly used necks.

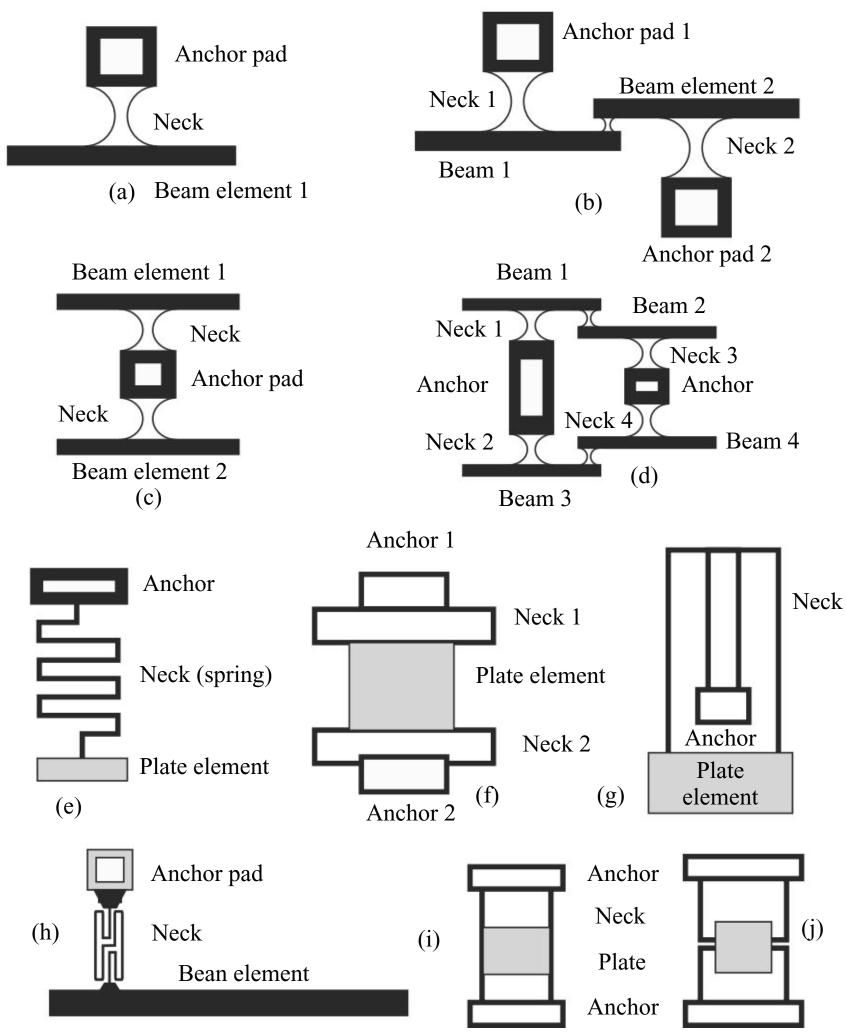


Fig. 4.23 *Various types of necks and flexure elements (a) Simple flexure with beam element and lever neck, (b) Cascaded flexure with beam element and lever neck, (c) Simple H-shaped flexure with beam element and lever neck, (d) Cascaded H-shaped flexure with beam element and lever neck, (e) Plate element with spring neck, (f) Plate element with larger dimension double parallel bar neck, (g) Plate element with folded neck, (h) Beam element with H-shaped neck, (i) Plate element with same dimension double parallel bar neck, (j) Plate element with larger dimension double parallel bar neck interfaced at center.*

4.4.8 Four-bar Linkages

One of the most common type of linkage is the four-bar linkage (Fig. 4.24), so-called because it has four axes of rotation connected by four rigid linkages. The four-bar linkage is a versatile mechanism, the design of which has been used extensively. Four-bar linkage has the capability of mimicking rotation, oscillation, and translation. Such mechanisms have a predictable and repeatable relative motion without any sliding or rolling contacts, resulting in a lack of friction and wear. The kinematics of a linkage is determined entirely by the shapes, sizes an material properties. Very complex input–output relationships can be realized with this mechanism. A variety of useful mechanisms can be formed from a four-bar mechanism through slight variations of proportions of links. Many complex link mechanisms are combinations of such mechanisms. Macro mechanism of this type are used in various applications, including control of optics, hammer guide springs, and measuring instruments.

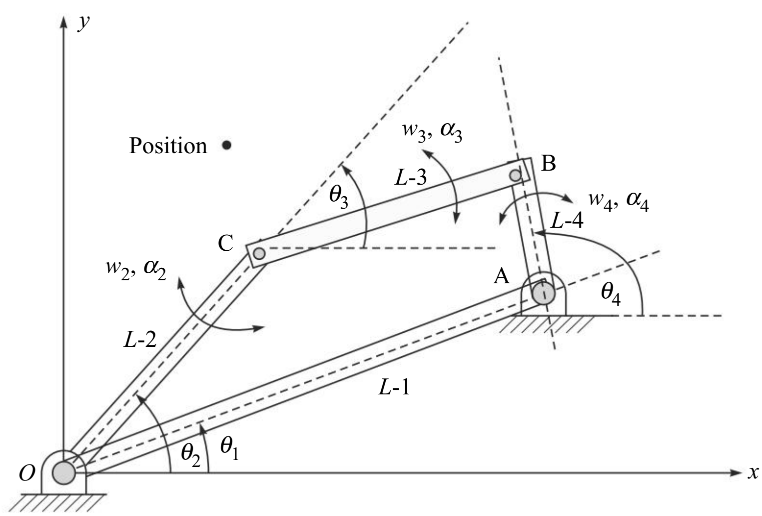


Fig. 4.24 *A typical four-bar linkage*

4.5 SUMMARY

This chapter started with basic need of a signal conditioning with its various circuits and systems used to achieve this. These systems are referred to as passive electronic systems (PES). Here, the rectification, clipping and clamping, amplification, filtering, isolation, bridge balancing comparison, etc. were discussed to understand the basic application of those in MEMS signal conditioning field. Also, the amplifier section including multi-stage instrumentation amplifier gives the essence of wide utility. The filter circuits as a frequency selector are very useful and were also discussed. Further, in addition ADC and DAC were also discussed. In addition, different passive mechanical systems (PMS) used in MEMS devices were discussed along with their specific application and classification. In particular, bearings, gears, rack and pinion, ratchet and pawl, slider and crank and four-bar linkage mechanisms were presented in proper order. Moreover, the chapter also describes the role of flexure systems in the MEMS devices.

Points to Remember

- MEMS device is composed of two main groups of functional systems; active and passive systems. Sensors and actuators are mostly considered as active systems.
- Active components are the important functional building blocks around which passive components are interfaced in order to take electronic or mechanical advantage.
- Passive systems are those which when interfaced with the active system constitute a full fledged MEMS device. Passive elements are the components, which typically do not directly involve in transducing power, rather they help in handling, manipulating and transferring power to a suitable form so that it will be easier to deal with.
- Passive systems are of two types; Passive Electronic Systems (PES) and Passive Mechanical Systems (PMS).
- The sole purpose for PES is to optimally modify some signal to a form that will be suitable to use. The PES *conditions* the available output signal of the transducer.
- *Signal conditioning* is a process of preparing the available signal into a desirable one.
- The unidirectional flow action is the basis of *rectification*, a process by which an alternating current (AC) is transformed or rectified into a direct current (DC).
- A *clipping* circuit transmits an arbitrary signal within the limits. The signal above and below the limits are suppressed. Clipping circuits are also referred to as voltage limiters.
- *Clamping* circuit shifts the input voltage levels without changing the shape of the signal.
- A circuit that accepts an input signal and produces an output signal in the same way as the input but has a larger magnitude is known as an amplifier circuit.
- When the filters are designed using only passive elements such as capacitors and inductors, they are referred to as *passive filters*.
- Filters that are designed either using a transistor or OPAMP are called *active filters*.
- The gain of the isolator or unity amplifier is unity.
- Zero crossing detector (ZCD) is a kind of comparator that provides a signal (pulse or step) at the output when the input signal passes the zero level.
- Oscillators are positive feedback circuits.
- An S/H circuit is one that samples the analog signal at a particular instant and retains the value for a specified time for subsequent use.
- A raw analog signal is converted into a series of samples by the process of sampling.
- Quantisation is a process of normalizing the amplitudes of the samples to its nearest *quantisation* or *discrete levels*.
- Aliasing cases the problem of overlapping of signal in the frequency domain.
- Mainly two techniques are employed as far as Digital to Analog Conversion is concerned, namely *binary resister-based* and *R-2R ladder-based* D/A converters, respectively. The latter one is better in terms of reliability.
- Mechanism, the heart of every mechanical construction is employed to take *mechanical advantage* and transmission of power and motion from one place to another.
- The bearing is a component that reduces frictional losses as surfaces side past one another. In magnetic bearings, mechanical friction losses are eliminated.
- Gears can help to increase or decrease force, torque or speed and can change the direction of the axis of rotation.
- More than two gears working together are called a *gear train*.
- A small electric motor capable of rotating at high speed may be providing enough power, but not enough torque.

- With a gear reduction, the output speed can be reduced while the torque is increased.
- A gear ratio is the relationship between how often the drive shaft turns to complete one full turn of the other or vice versa.
- The velocity ratio of the driver and follower does not change by putting any number of gears between them.
- The *slider–crank* is a linkage that transforms linear motion to circular motion or vice versa.
- In a vehicle the IC engine transforms the reciprocating motion of a piston to a rotary motion at the crank-shaft.
- Geneva wheel based mechanism is used for achieving intermittent motions.
- Four-bar linkages have the capability of mimicking rotation, oscillation, and translation.
- Welding and nut-bolt joint methods are not applicable to MEMS systems and devices.
- In many occasions, the immovable elements are not directly attached to the anchors, rather they are interfaced through spring or special cantilever arrangement.
- The design of flexure primarily looks at the bending moments in relation to geometry of the neck.
- Practically the flexure utilizes thin necks. There are several ways to anchor the microelement (active element) through the neck.

Exercises

1. What do you mean by Passive Mechanical Systems (PES) and Passive Electronic Systems? Give some examples.
2. With a simple schematic diagram show the concept of System-on-a-chip (SoC) design method. Why is SoC implementation principle important in MEMS design? Give some example of application of futuristic SoC devices.
3. Briefly define the principles and applications of the following Passive Electronic Systems (PES).
   (a) Rectification
   (b) Clipping and clamping circuits
   (c) Amplifier
   (d) Isolator
   (e) Comparator
   (f) Oscillator
   (g) Sample and hold circuits
4. What are the different types of filter circuits you know? What does the frequency response curve of a typical filter signify? Why are upper cut-off and lower cut-off frequencies important? Define bandwidth of a typical bandpass filter. What is the value of power at these frequencies? Discuss differences between active filter and passive filter.
5. Why is instrumentation amplifier significant? How is it realized? Explain with a suitable diagram.
6. "Wheatstone bridge is not a transducer"—Justify. Why is one of the arms of the Wheatstone bridge circuit called active arm? How is the transducer connected? Why is instrumentation amplifier preferred to interface at the output of the bridge? Explain the principle of operation of a Wheatstone bridge circuit.
7. What are the different methods of converting analog signal to its digital equivalent? Draw the schematic diagrams of various types of analog to digital converter (ADC) circuits and explain their principles of operations. Discuss the relative merits and demerits of these methods.
8. Define the following terms with respect to ADC.
   (a) Resolution
   (b) Accuracy
   (c) Sampling rate and aliasing

9. What are the different methods of converting digital signal to its analog equivalent? Draw the schematic diagrams of various types of digital to analog converter (DAC) circuits and explain their principles of operations. Discuss the relative merits and demerits of these methods.
10. Define the principles and applications of the following Passive Mechanical Systems (PMS).
    (a) Bearings
    (b) Rack and pinion
    (c) Ratchet, pawl and crank
    (d) Slider and crank
    (e) Geneva wheel
    (f) Four-bar linkage
11. What do you mean by mechanical advantage? In particular, how is this achieved by gearing mechanism? Distinguish between the gearing-up and gearing-down mechanism. Name the gears that are present in a gear train. Which type of gearing mechanism is called torque amplifier? Write down the mathematical expression of the gear ratio. "The velocity ratio of the driver and follower does not change by putting any numbers of gears between them"—Justify.
12. Discuss various ways of designing neck and flexures for the MEMS devices. Draw their diagrams. Why are flexures called PMS?

Chapter 5

Mechanical Sensors and Actuators

Objectives

The objective of this chapter is to study the following.

- To elucidate the role of microsensors and microactuator with examples
- To discuss the principle of sensing and actuation
- To list out the types of measurands to be measured
- Principle of various types of mechanical microsensors and microactuators
- Cantilever, beam and diaphragm type microstructures as active element
- Principle of operation of flow sensor and pressure sensor
- Capacitive effects in sensing and actuating
- Application of MEMS microphones for acoustics measurements
- MEMS gyroscope as rate sensor
- Principle of shear mode piezoactuator
- Typical gripping piezoactuator
- Inchworm Technology and actuation principle

5.1 INTRODUCTION

Sensors and actuators play a significant role in almost all kinds of instrumentation, process control, factory automation and machine control applications. For simple understanding consider the machine control applications such as the mechatronic systems. A mechatronic system is an electronically controlled mechanical machine. Some examples of mechatronic systems are robot, AGV (Autonomous Guided Vehicle), cranes, NC (Numerically Controlled) machines, automobile engine and tanker. Micromechatronic systems are mechatronic systems but their dimensions are measured in microscale levels. Regardless of their dimension all mechatronic systems are composed of four prime components. They are sensors, actuators, controllers and mechanical components. Figure 5.1(a) shows a schematic diagram of such systems comprising all the components stated above. The schematic illustration is common to all types of mechatronic systems[1]. Figure 5.1(b) shows an example of motion control

[1] For more information on mechatronic system refer to Chapter 1 of the Book *Mechatronic: Principles, Concepts and Applications*, by Mahalik published by McGraw-Hill, Singapore.

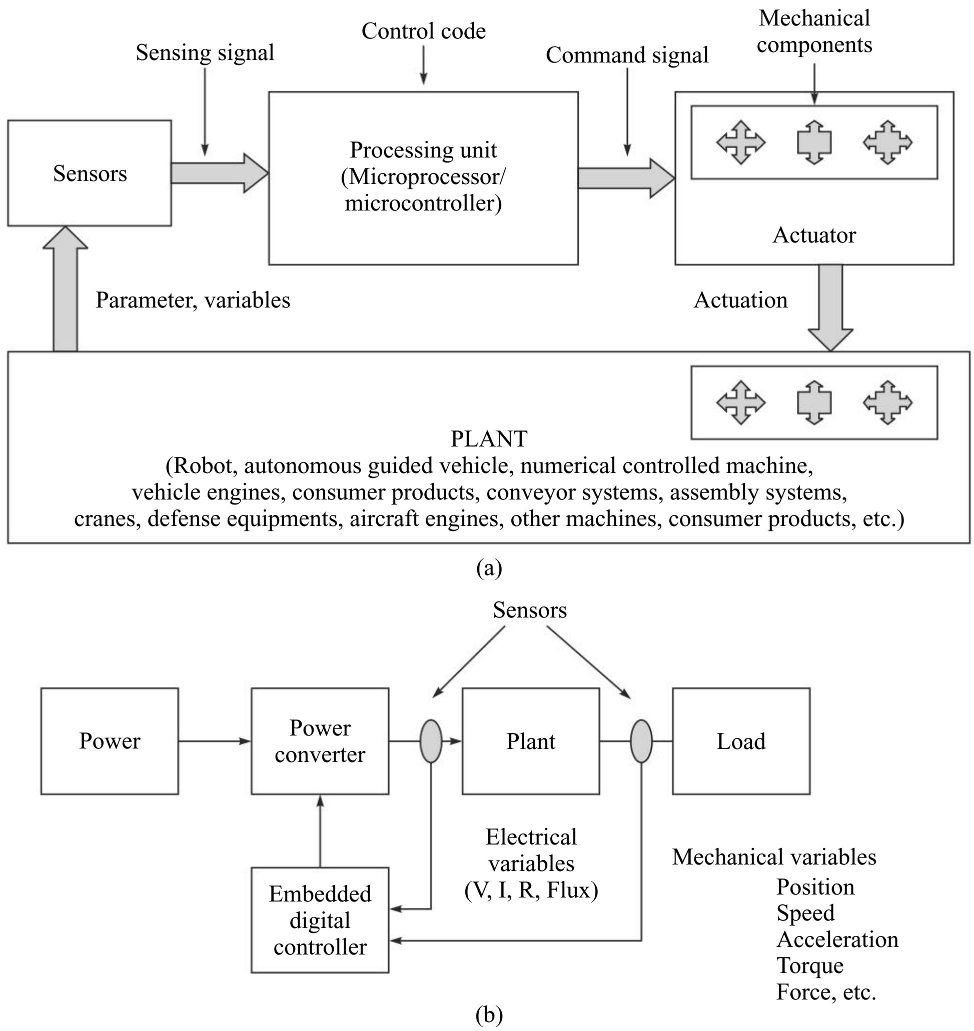


Fig. 5.1 *(a) A simple sensor-actuator connection, (b) Typical example with embedded motion control system (The feedback system in this above motion control system can use MEMS sensors)*

system. The plant could be an electrical motor. The meaning of load is simply the mass of the rotor along with the shaft which needs to be rotated at a precise predefined speed. The output speed can be observed by using the feedback sensors. Detail discussion on feedback control is beyond the scope of this chapter. Students are advised to read the book mentioned in the footnote on previous page.

Although, the design and construction of macroscale and microscale mechatronic system vary but the principle behind the control interfacings remain similar.

From control point of view sensor will sense the signal and actuator will activate a control action. Depending upon the control requirements sensors are chosen. The role of controller is to execute the *control code* for a given task. In its abstract form, a simple but exemplar control code could be;

'Drive the actuator from position 1 to position 2 when event-n occurred'

The above piece of control code (also called control algorithm) can be realized in a variety of ways through a broad range of sensors, processing unit (the physical controller), and actuating systems. The role of sensor in this case is to check whether or not the *event-n* (could be a switching operation detecting the presence of an object) has occurred. If yes, it passes the information to a controller. On receipt of this information controller generates necessary driving signal in order to drive the actuator so that the end-effector of the actuator can be moved from position 1 to the position 2. The act of observing the temporal occurrence of physical phenomenon is called *sensing*, whereas the act of driving an end-effector is called actuation.

5.1.1 Sensing

Sensing is performed by using transducers. For an example, photodevices, used to measure the intensity of light are considered as transducers. Transducer is defined as a basic element that converts or *transforms* one form of energy to another form, usually to electrical energy. The question is why is it necessary to convert one form of energy to another form and what are those forms? The immediate answer to this question is very clear in the sense that the signal or energy in the current form in fact is not suitable (sometimes not possible) for observation, measurement, quantification or manipulation.

Transducers are mainly used to quantify the physical, electrical, fluidic and mechanical *variables* such as temperature, pressure, magnetic field, voltage, flow, vibration, radiation, and so on. Such parameters can commonly be called as *measurands*. Some of the important measurands are listed as follows.

- Acoustic field and noise
- Acceleration
- Density
- Force and load
- Humidity level
- Length
- Magnetic fields
- Phase
- Resistance
- Speed
- Vibration
- Amplitude
- Color
- DNA and Protein
- Frequency
- Image
- Level
- Pressure
- Power
- Rotation
- Temperature
- Vacuum
- Angle
- Current
- Flow
- Gas concentration and pH
- Intensity of light
- Motion
- Position
- Radiation
- Stress and strain
- Velocity
- Voltage

For the observation, measurement, quantification and/or manipulation of measurands, different types of transducers with different *transduction principles* are used. For example, the transducer that is used for the detection of temperature is different from a transducer that is employed for the detection of ultrasonic sound signal. (The term "ultrasonic" applied to sound above the frequencies of audible sound, and nominally over 20 kHz. Signal used for medical diagnostic ultrasound scans extend to 10 MHz and beyond.) For various applications, a variety of transducers, with different transduction principles are designed and fabricated.

The term transducer and sensor have been used synonymously although the concepts are little different in the macro level system design. Transducers are the physical element, which is a part of a sensor. Indeed, transducer is an essential element of a sensor. A sensor is merely a sophisticated transducer because it contains the *signal conditioning circuits* capable of amplifying and refining the weak and raw signal that is available at the output of the transducer. Some of the commonly used signal

conditioning circuits are amplifiers, noise filters, offset compensator, and Analog to Digital Converter (ADC). Figure 5.2 gives an illustration of a sensor. The input signal refers to as measurand. The output of the transducer is referred to as *equivalence*. The signal/energy at the output of transducer and hence at the output of the sensor is in different form with respect to the input signal/energy. In this book transducers and sensors will be used synonymously, because in the study of microsystems, they are treated equally.

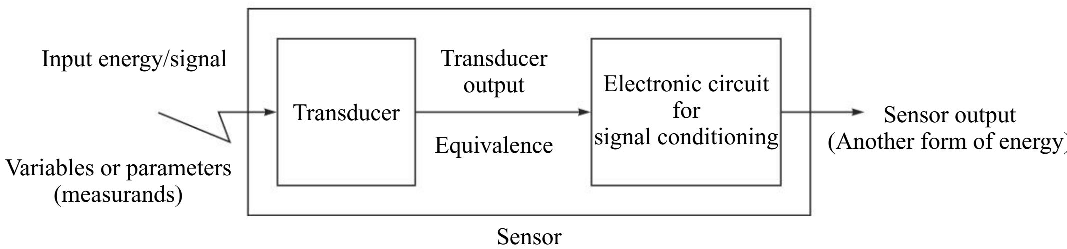


Fig. 5.2 *A typical sensor showing transducer and signal conditioning unit*

Depending upon the application requirements various types of mechanical microsensors are fabricated. This chapter describes the principle of operation of the important mechanical microsensors developed in the MEMS domain. Sensing is possible through micromechanical components adopting microactuation principle! At this stage it is sufficient to say that some microactuators can also be used as microsensors in some other applications. For instance, when a microcantilever (microactuator) tip, which is under deformation due to controlled bias signal, touches another element in a system, a kind of touch signal is said to be generated and is readout. In another example, the phenomenon of piezoelectricity and piezomechanics of piezoelectric materials can be exploited to fabricate the microactuators as well as microsensors.

5.1.2 Actuation

Actuation is simply the process of conversion of one form of energy to mechanical form. The process is known as the principle of energy transduction. A device that accomplishes the conversion is termed actuator. Actuators play a very important role in automation and control. Various types of actuators are used depending upon many factors including force, torque, speed of operation, accuracy, precision, power consumption and performace.

As already indicated, the basis of actuation is the method of energy transduction. Principal energy transduction methods, called driving methods, are mechanical, thermal, electrical and magnetic method. Accordingly the commonly available microactuators are called:

- Mechanical actuators
- Thermal actuators
- Electrostatic actuators
- Magnetic actuators

Mechanical actuators are broadly categorized under two groups, namely (i) Mechanical structure based microactuator and (ii) Active material based microactuators. There are many structures of microactuators, in which energy transformation takes place. In view of this, the mechanical structures based microactuators are beam type, cantilever type, plate type or membrane type. The plate or membrane type microactuators are also called diaphragm type microactuators. The spatial geometries of these mechanical structures change when control or bias signals are applied to the components,

validating the principle of actuation. The structures go to their original shapes when the control signals are withdrawn. So the principle of reversibility holds good.

Materials, which undergo some sort of transformations through physical interactions, are referred to as *active materials*. The principle of operation of these actuators is based on the fact that they make use of properties of the material from which they have been constructed. Some of the important active materials are *piezoelectric* material and *magnetostrictive material*. Piezoelectric actuators use piezoelectric materials. Application of a voltage to a piezoelectric material causes a small motion. Magnetostrictive material undergoes a change in shape either due to change in the phase or magnetization state of the material. Phase change occurs by heat flux whereas state change occurs because of magnetic field. Magnetostrictive materials such as shape memory alloys (SMA) like Nitinol react to heat, whereas Ferromagnetic Shape Memory Alloy (FSMA) materials, such as *Terfenol-D* (Tb0.3Dy0.7Fe2), an example of another magnetostrictive material, changes its shape when it is subjected to magnetic field. The actuator built with SMA is called *memory metal actuator*. As material researchers have focused on improving the performance of the different families of active materials during the past few years, engineers have been developing ways to use them in designing practical actuating devices. The potential applications of microactuators are:

- Aerospace structure monitoring systems
- Automotive monitoring and control devices
- Fluid control devices and biomedical equipment
- Precision optical components manufacturing and handling
- Precision manufacturing and process-monitoring equipments
- Microassembly

This chapter describes MEM components as far as mechanical sensors and actuators are concerned. In particular, the chapter will be dedicated to describe the concept, principle and applications of mechanical sensors and actuators. The structural design and their impact on sensing and actuation will be discussed. The important discussion includes principle of operation, design methodology, drive mechanism and control methods of:

- Pressure sensor (MEMS microphone)
- Flow sensor
- Microgripper
- Rate sensor (MEMS Gyroscope)
- Inchworm actuator

In the following chapters thermal, electrostatic and magnetic actuators will be discussed elaborately.

5.2 PRINCIPLES OF SENSING AND ACTUATION

This section describes the principle of sensing and actuation of mechanical MEMS. Mechanical MEMS can be a microsensor or a microactuator. Mostly, cantilever, beam, and diaphragm structures form the basis of the mechanical MEMS. Piezoelectric materials are used as the active materials. Piezoelectricity concerns with mechanical properties such as stress and strain. This is the reason why piezoelectric material based MEMS devices have been classified under mechanical MEMS. The mechanical MEMS utilizes the following methods and principles.

- Beam and cantilever methods
- Capacitive method
- Piezoelectric principle

5.3 BEAM AND CANTILEVER

Micro beams and cantilever structures are essentially very useful transducers elements using which many physical phenomena can be measured. The principle behind this lies with the deflection of the beam and cantilever structures. The deflections are picked up either by capacitive or piezoresistive measurement. The difference between a beam and cantilever is that beam is fixed at both the ends whereas a cantilever is fixed at only one end.

Figure 5.3 shows various types of beams and cantilevers and their operational modes. The beam shown in Fig. 5.3(a) is at the resting state and therefore initially it is horizontal and straight. When force is applied the horizontal axis of the beam is deformed into a curve. We let x denote the horizontal distance of that a point from the left-hand end of the beam, and we define $d(x)$ to be the vertical displacement of the beam. The graph of the function 'd' is called the *deflection curve* for the beam. The deflection of a beam depends on its length, its cross-sectional shape, the material, the point at which the deflecting force is applied, and how the beam is supported. Figure 5.3(b) is a two-layer deflection beam, which is hinged at the center. This type of two-layered beam is used for bi-stable applications such as relay. They have two stable states; up-state and down-state.

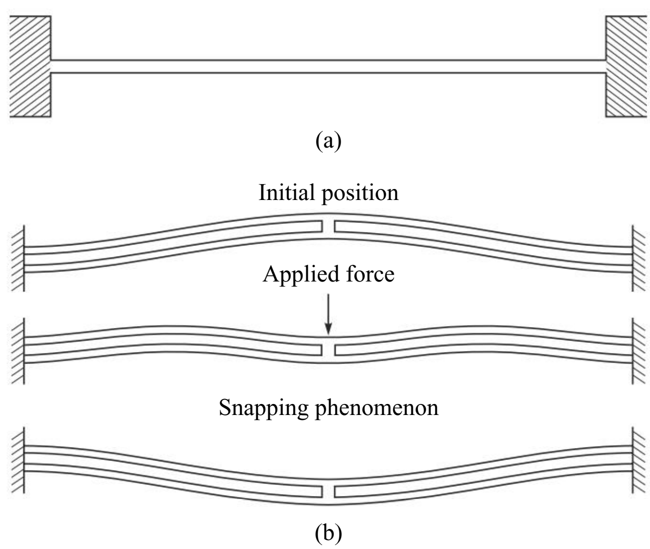


Fig. 5.3 *(a) A schematic of a single layer microbeam; (b) A schematic of a double layer microbeam hinged at middle*

Figure 5.3(c–e) and Figure 5.3(f–h) are called cantilevers and beams, respectively. One of the end is free in case of cantilever whereas both the ends are supported in case of beams as shown. Load or force can be applied at the preferable point usually in between the central point and the free end. However, other configurations are in many cases requisite. The load can also be distributed, but rarely seen in the MEMS devices. The deflections and the slopes are illustrated. The normal and maximum deflection ($d(x)$, d_{max}) and slope ($\theta(x)$, θ_{max}) formulae are given in Table 5.1. The maximum deflection of various beams can be obtained by putting the value of x as appropriate. Where, P is the applied force, E is the modulus of elasticity, J is called centroidal moment of inertia of the cantilever/beam cross-section, L is the length of the cantilever/beam, a and x are denoted as shown in the figures. The governing equation

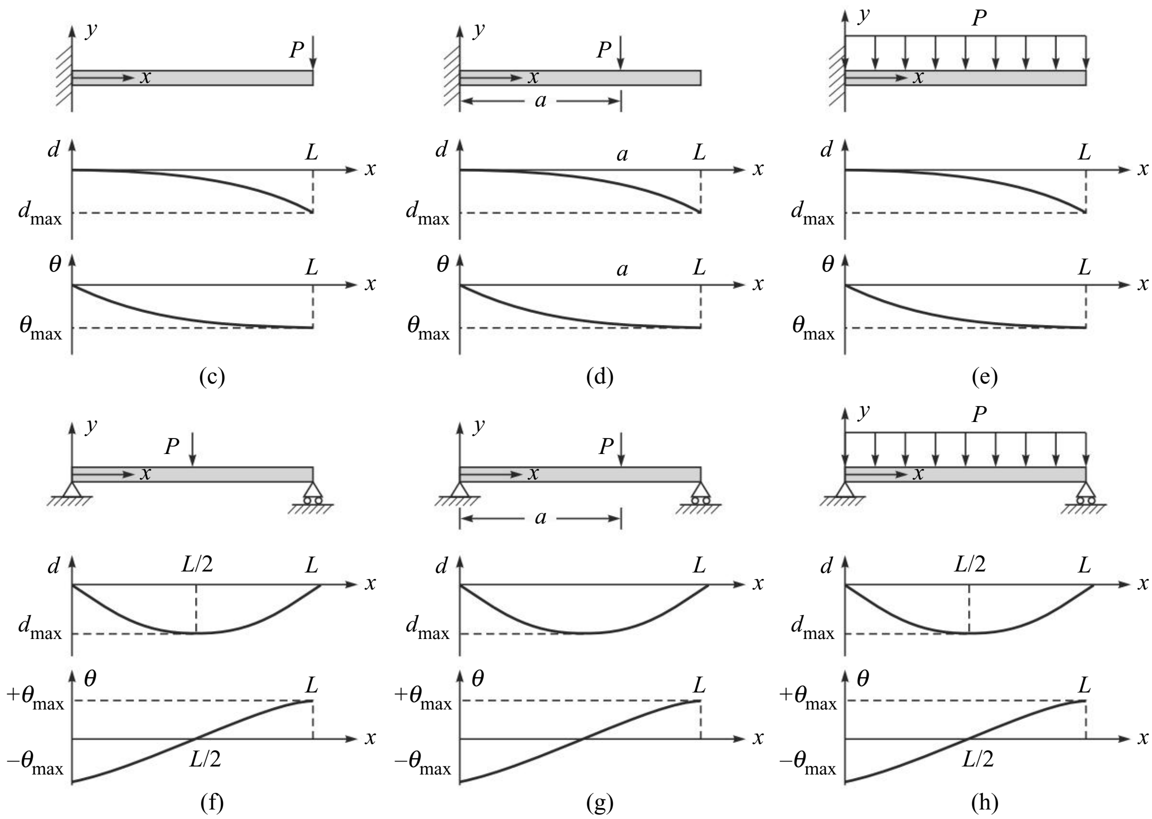


Fig. 5.3 *(c) Point load at the end, (d) Point load at the intermediate point, (e) Point load with Uniform Distributed Load, (f) Simple supported beam with point load at the middle, (g) Simple supported beam with point load at the intermittent point, (h) Simple supported beam with uniform load*

with regard to deflection is based on the way the beam is supported, which is captured by the boundary conditions of the differential equation.

5.4 MICROPLATES

Various structures and features of MEMS plates are square, rectangular, solid, porous, single-layered and multi-layered. Depending upon the application requirements they are designed and subsequently used. The effective area of the plates ranges from few hundred to thousand μm^2. One application of microplates is to build parallel plate capacitors. In particular, a voltage tunable parallel plate capacitor has two-movable parallel plates. Using electrostatic method one can move the plates. Typically, capacitance of the tunable capacitor can vary from 1.0 pF to 2.0 pF corresponding to applied bias voltage from 0.5 V to 3.0 V.

5.4.1 Diaphragm Structures and Theory

At macro level, diaphragms are thin circular plates broadly used for the measurement of both low and high values of pressure, force or load. The principle is based on deflection. The displacement '*d*' is

Table 5.1 *Deflection equations*

Type of applied force	*Deflection*	*Slope*
Point load at the end	$d(x) = -\frac{Px^2(3L-x)}{6EJ}$ $d_{\max} = d(L) = -\frac{PL^3}{3EJ}$	$\theta(x) = -\frac{P(2L-x)\,x}{2EJ}$ $\theta_{\max} = \theta(L) = -\frac{PL^2}{2EJ}$
Point load at the intermediate point	$d(x) = -\frac{Px^2}{6EJ}(3a-x) \quad 0 \le x \le a$ $-\frac{Pa^2}{6EJ}(3x-a) \quad a \le x \le L$ $d_{\max} = d(L) = -\frac{Pa^2}{6EJ}(3L-a)$	$\theta(x) = -\frac{Px}{2EJ}(2a-x) \quad 0 \le x \le a$ $-\frac{Pa^2}{2EJ} \quad a \le x \le L$ $\theta_{\max} = \theta(x > a) = -\frac{Pa^2}{2EJ}$
Point load with uniform distributed load	$d(x) = -\frac{px^2(6L^2-4xL+x^2)}{24EJ}$ $d_{\max} = d(L) = -\frac{pL^2}{8EJ}$	$\theta(x) = -\frac{px(3L^2-3xL+x^2)}{6EJ}$ $\theta_{\max} = \theta(L) = -\frac{pL^3}{6EJ}$
Simply supported beam with point load at the middle	$d(x) = \begin{cases} -\frac{Px(3L^2-4x^2)}{48EJ} & 0 \le x \le L/2 \\ \frac{P(L-x)(L^2-8xL+4x^2)}{48EJ} & L/2 \le x \le L \end{cases}$ $d_{\max} = d(L/2) = -\frac{pL^3}{48EJ}$	$\theta(x) = -\begin{cases} -\frac{P(L^2-4x^2)}{16EJ} & 0 \le x \le L/2 \\ \frac{P(3L-2x)(2x-L)}{16EJ} & L/2 \le x \le L \end{cases}$ $\theta_{\max} = \theta(0\,\&\,L) = \mp\frac{PL^2}{16EJ}$

Contd.

Table 5.1 *Contd.*

Type of applied force	*Deflection*	*Slope*
Simply supported beam with point load at the intermittent point	$d(x) = \begin{cases} -\dfrac{(L-a)\,Px(-a^2+2La-x^2)}{6EJL} & 0 \le x \le a \\ \dfrac{aP(L-x)(a^2+x^2-2Lx)}{6EJL} & a \le x \le L \end{cases}$ $d_{\max} = \begin{cases} -\dfrac{a^{3/2}P(L-a)(2L-a)^{3/2}}{9\sqrt{3}EJL} & 0 \ge L/2 \\ \dfrac{aP(L^2-a^2)^{3/2}}{9\sqrt{3}EJL} & a \le L/2 \end{cases}$	$\theta(x) = \begin{cases} -\dfrac{P(a-L)\,(a^2-2La+3x^2)}{6EJL} & 0 \le x \le a \\ \dfrac{aP(a^2+2L^2+3x^2-6Lx)}{6EJL} & a \le x \le L \end{cases}$ $\theta_{\max} = \begin{cases} \theta(0) = -\dfrac{aP(a-2L)(a-L)}{6EJL} & a \le L/2 \\ \theta(L) = \dfrac{aP(a^2-L^2)^{3/2}}{6EJL} & a \ge L/2 \end{cases}$
Simply supported beam with uniform load	$d(x) = -\dfrac{px(L^3-2x^2L+x^3)}{24EJ}$ $d_{\max} = d(L/2) = -\dfrac{5pL^4}{384EJ}$	$\theta(x) = -\dfrac{p(L^3-6x^2L+x^3)}{24EJ}$ $\theta_{\max} = \theta(0\ \&\ L) = \mp\dfrac{pL^3}{24EJ}$

(*Source:* http://www.efunda.com)

proportional to the applied pressure, force or load (Fig. 5.4). The pneumatic, fluid and acoustic applications use diaphragms. The mathematical deflection formula can be written as follows:

$$d = \frac{ca^2 F}{t^3} = kF \qquad (5.1)$$

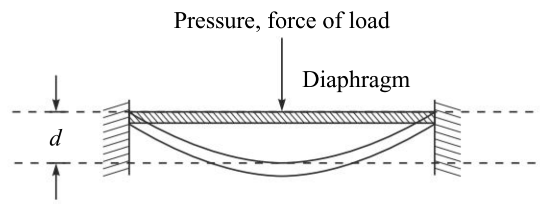


Fig. 5.4 *Diaphragm*

Here, $k = ca^2/t^3$; F is the applied force, a and t are radius and thickness of the diaphragm, respectively, c is a constant which depends on Poisson's ratio and Young's modulus of the material, from which the diaphragm has been made up.

k in Eq. 5.1 can be defined as sensitivity of the transducer. The above equation can be written more precisely. The deflection then can become,

$$d(r) = \frac{Pr^4}{64\,k}\left\{1 - \left(\frac{x_r}{r}\right)^2\right\}^2 \qquad (5.2)$$

where P is the applied load, x_r is radial coordinate, r is diaphragm radius, k is a measurement of stiffness, given by

$$k = \frac{Et^3}{12(1-v^2)} \qquad (5.3)$$

where E, t, and v are Young's modulus, plate thickness, and Poisson's ratio, respectively.

5.4.2 Corrugated Diaphragms

In order to improve the sensitivity, *corrugated diaphragms*, shown in Fig. 5.5, are designed. These are called *capsules* and *bellows*. Bellows have more sensitivity than the capsules, although their application domains vary. The materials used for diaphragms are nickel, phosphor and stainless steel. In many pressure sensors, strain gauges are mounted on the diaphragm. These are called strain gauge based electrical output pressure transducers. The change in pressure is related to the strain in the diaphragm. Sometimes the diaphragm itself is made up of semiconductor and the strain gauge is doped within the diaphragm at the time of manufacturing. Such types of transducers are intended for the measurement of

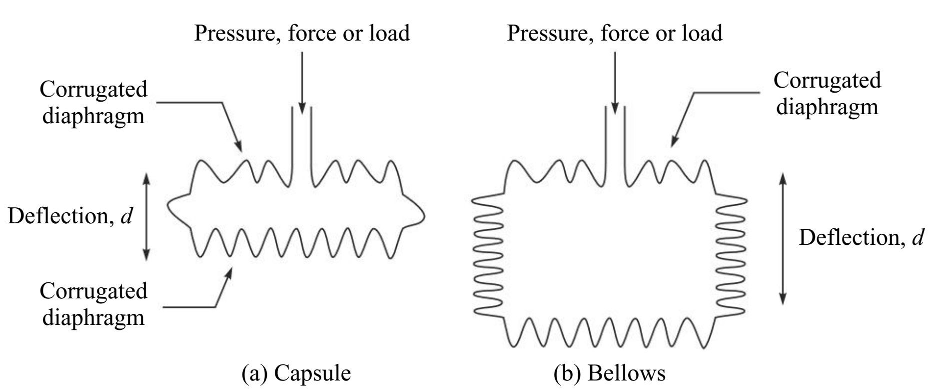


Fig. 5.5 *Capsule and bellows for pressure measurement*

absolute pressure (pressure with respect to vacuum). Capsules and bellows are broadly macro level devices! Microlevel designs are also possible.

5.4.3 Deflection Formula for MEMS Diaphragm

The design criteria (thickness, radius, etc.), fabrication methods (material selection, surface, bulk, etc.) and application areas (absolute pressure, acoustic pressure, etc.) of MEMS diaphragms significantly differ although the physical principle remains same, i.e. deflection is proportional to the measurand. Quantitatively, the deflection theory and characteristics of micromachined diaphragms is not same as that of large deflection theory and characteristics. For micromachined diaphragms it is assumed that the applied load is constant over the diaphragm surface. In contrast to the deflection equation mentioned in Eq. 5.3, the deflection in the MEMS diaphragm is, however, given by,

$$d(r) = \frac{Pr^2}{4\sigma_i t}\left\{1 - \left(\frac{x_r}{r}\right)^2\right\} \tag{5.4}$$

where σ_i is the intrinsic built-in stress of the plate (diaphragm).

The diaphragms are manufactured in various shapes and sizes depending upon the nature of transduction signal at hand. Most commonly diaphragm shapes are either circular or rectangular, however the thickness in a particular type is constant and usually less than 10 μm. The development of high-performance microdiaphragm structure is critical in the micromanufacturing process. The diaphragms should provide linear relationship with regard to deflection and applied pressure. Making diaphragm thinner can increase the sensitivity. But for high-pressure applications, the thin diaphragm produces nonlinear responses. It is important to characterize the relationship between diaphragm thickness, deflection, and sensitivity, both analytically and experimentally prior to its design for real applications. The analytical analysis primarily provides appropriate guidelines for the design of microdiaphragm-based sensors.

5.4.4 Piezoresistive Measurement

CMOS-based ultrasonic MEMS diaphragm is used for fault detection: by sensing the echo of an ultrasonic burst in a structure. In such application acoustic impedance of the diaphragm is exploited. Acoustic impedance is defined as the ability of a diaphragm material to transmit sound. The diaphragm is integrated with other elements such as piezoresistor, which is built in CMOS polysilicon layer. The piezoresistor behaves as a strain gauge through which piezoresistive sensing is achieved. The change in resistance is taken as read out and is produced through a balanced bridge circuit (Wheatstone bridge). The bridge circuit can also be fabricated in the same substrate. The output of the bridge circuit is fed to the instrumentation amplifier. One can note that acoustic impedance of the diaphragm is always less than the medium, so that the piezoresistive-implanted strain gauge diaphragm will be able to closely track a displacement due to a sound wave. Optical and capacitive based sensing mechanism can also be employed in the diaphragm sensors.

5.4.5 Free-standing Diaphragm

There is another structure of diaphragms called free-standing diaphragms which possesses high sensitivity (i.e. the ratio of output to the input is high). Fabrication of free-standing diaphragms (Fig. 5.6) is required for the design of micromirrors. Micromirrors have application in display technology and optical switches in the telecommunications industry and consumer product arena (See Chapter 7).

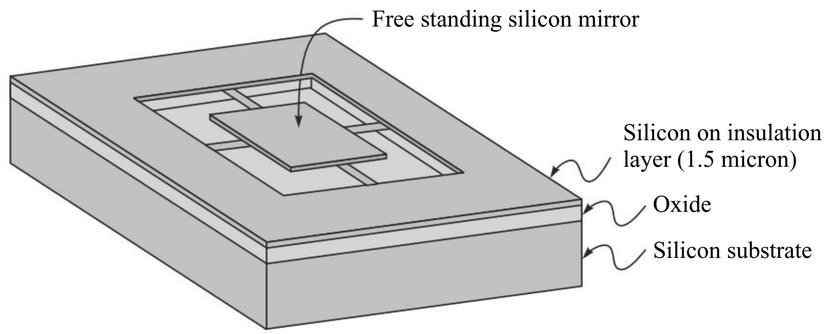


Fig. 5.6 *Free standing diaphragms*
Source: Graff and Schubert, *Sensors and Actuators,* 84, 2000, Elsevier

Moreover, these devices find their applications as tunable filters, microcavity and wavelength-sensitive detectors. Using a wet chemical etch and a release process, free-standing diaphragm structures can be fabricated. Electrostatic actuation to the free-standing diaphragm is possible for tuning applications in optical transmission systems because of the formation of a cavity by the membrane and the substrate surface.

5.5 CAPACITIVE EFFECTS

Basically, the geometrical dimensions as well as the property of the dielectric material of a parallel plate type capacitor are exploited to quantify many types of measurands. The principle of capacitive transducer is based on the change of *distance*, *area* and/or the *permittivity*. The basic equation governing the principle of operation is expressed as follows.

$$C = \frac{\varepsilon_r \varepsilon_0 A}{d} \tag{5.5}$$

where, d is the distance, A is the area, ε_r is the relative *permittivity* of the dielectric material exists between the plates and ε_0 is the permittivity of the free space (The value is 8.854*10E-12 Farad/Meter). The distance and overlapping area can be changed either by the movement of one of the plate relative to other. Displacing the dielectric material can also vary the capacitance. Figure 5.7 illustrates three ways of measuring the measurands. In Fig. 5.7(a) the distance between the plates has been displaced in accordance with the input measurand. Figure 5.7(b) and 5.7(c) shows different situations, where one of the plate and the dielectric material have been displaced in accordance with the measurand, respectively.

5.5.1 Capacitive Sensors

Capacitive pick-ups can be used as a proximity sensor. They are designed to operate by generating an electrostatic field and detecting changes in this field caused when a target or object approaches the sensing area. The sensor consists of various units such as a capacitive probe, an oscillator, an amplifier, a signal rectifier, a filter circuit and an output circuit. Usually the amplifier, rectifier and filter are integrated in one unit. In the absence of a target, the oscillator is inactive. As a target or object to be detected approaches, it raises the capacitance of the probe system due to the reason that the distance between the target and sensor face decreases. The face of the sensor and target can be thought of as a capacitor. When the capacitance reaches a specified threshold value, the oscillator gets activated. The

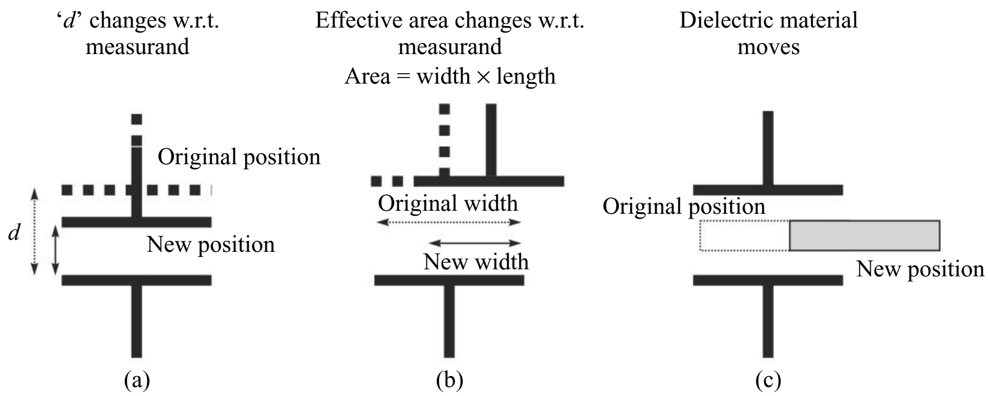


Fig. 5.7 *Principle of operation of capacitive transducers*

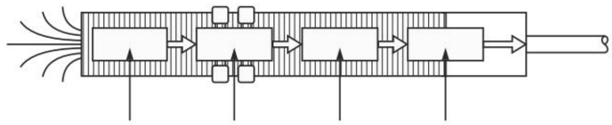
output of the oscillator is amplified in order to increase the signal strength. The oscillating signal is then rectified suitably to trigger the output circuit to change between on and off. One can note that the capacitance of such probe system is determined by the target size, dielectric constant and distance from the probe. Bear in mind that the larger the size and dielectric constant of a target, the more it increases capacitance.

Capacitive sensors are easily fabricated by surface micromachining process. On the top of the substrate an electrode layer is deposited. Depositing a thin film again on top of the electrode layer a capacitive MEM sensor is produced. Between the electrode layer and thin film, a gap has to be created. The surface micromachined sensors mostly use the capacitive transduction method to convert the input mechanical signal to the equivalent electrical signal. When in use, the applied physical stimulus can cause one of the plates to move. The gap between the two electrodes changes causing a change in the capacitance. This change in capacitance is the electrical equivalent of the input mechanical stimulus. Such sensors are referred to as parallel plate mechanical sensor. Figure 5.8(b) shows a surface micromachined based mechanical capacitive sensor.

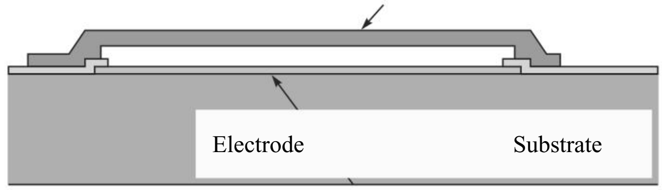
Capacitive sensors are non-contact type for which accuracy and resolution are high. However, because of appearance of parasitic capacitances from the cable interconnections and the environmental effects such as moisture, pressure, vibration and dust, the performance is likely to degrade. The manufacturers of capacitive sensors take care of these very important factors during micromachining and packing. The capacitive sensors are employed for the measurement of audio signal, acceleration, displacement, flow, force, level, pressure, relative humidity, strain, thickness and velocity.

5.6 PIEZOELECTRIC MATERIAL AS SENSING AND ACTUATING ELEMENTS

Most materials have physical properties, which are difficult to alter. However, there exist some materials whose physical property can be significantly altered. One such material is piezoelectric material. Piezoelectric materials are widely used as sensing and actuating element in various applications. This section describes two important properties of piezoelectric materials such as piezoelectricity and piezomechanics.



(a)



(b)

Fig. 5.8 *(a) Different parts of a capacitive pick-up based sensor, (b) Capacitive MEMS sensor*

5.6.1 Piezoelectricity

Piezoelectric materials are permanently polarized crystalline materials. If their dimensions are changed as a result of mechanical force, electric charges proportional to the imposed force are accumulated on the surface upon which the force is imposed (Fig. 5.9 and Fig. 5.10). The overall crystal remains electrically neutral, but displacement of charge center results. The phenomenon is known as the piezoelectric effect or *piezoelectricity*, which was originally studied by Pierre and Jacques Curie in 1880. Some of the piezoelectric materials are ammonium dihydrogen phosphate, barium titanate, lead betaniobate, lithium sulphate, polyvinylidine fluoride (PVDF), Rochelle salt, tormaline, quartz and zirconate titanate. The Greek word 'piezo' bears the meaning of pressure. The property is exploited to measure many physical parameters such as force, pressure, strain, torque, acceleration, sound and vibration. The relationship between the applied force and the charge accumulated is given by,

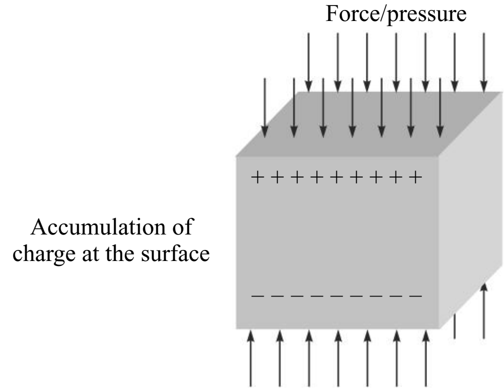


Fig. 5.9 *Piezoelectric effect (piezoelectricity)*

$$Q = cF \tag{5.6}$$

where, F is the applied force, Q is the accumulated charge and c is called the piezoelectric constant. The piezoelectric constant is not actually a constant; rather the value depends on the orientation and the mode of operation. Orientation and mode of operation refers to as follows.

Orientation	:	By which the transducers are cut from the original crystal.
Mode	:	The way by which the stress is applied.

Piezoelectric transducers are cut from the crystal slabs or plates in a specific orientation with respect to the crystal axes depending on the application requirement. There are mainly four types of cuts, namely;

- Longitudinal cut
- Transverse cut
- Shear cut
- Polystable cut

The shear cut is commonly used for force and acceleration measurement whereas *polystable* cut based transducers are employed for pressure measurement, in high temperature environment. Various modes of operation are bend, shear and compression. The charge in a piezoelectric transducer is therefore proportional to the piezoelectric constant of the material and to the stretched or squeezed force. During operation, the piezoelectric transducer can be thought of as a parallel plate capacitor, which is shown in Fig. 5.10.

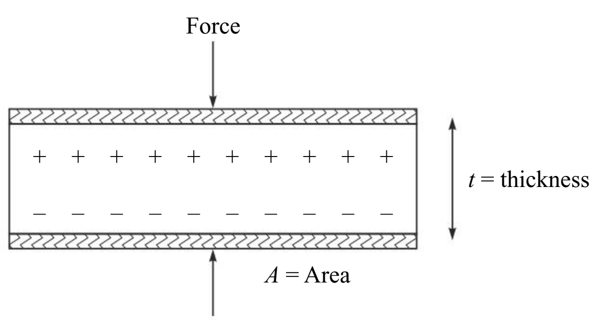


Fig. 5.10 *Piezoelectric capacitance*

The value of the capacitance is given by (see Eq. 5.5 also),

$$C_p = \frac{\varepsilon_r \varepsilon_0 A}{t} \tag{5.7}$$

where, C_p is called piezoelectric capacitance. A is the area, t is the thickness of the transducer, ε_r is the relative permittivity of the piezoelectric material and ε_0 is the permittivity of the free space (The value is 8.854*10E-12 farad/m). The capacitance can be expressed in terms of charge accumulated per unit voltage as expressed in Eq. 5.8.

$$C_p = \frac{Q}{V} \tag{5.8}$$

where V is the potential difference between the plates. Now from Eqs 5.6, 5.7, and 5.8 we can write,

$$V = kF \tag{5.9}$$

where
$$k = \frac{ct}{A\varepsilon_r \varepsilon_0}.$$

Thus, from Eq. 5.9, one can note that the voltage is proportional to the imposed force. The potential difference, V is very weak. Usually, the voltage is amplified by using an amplifier. The amplifier used for this purpose is referred to as *charge amplifier* due to the reasons that the accumulation of the charge effectively appeared as a difference in the potential across the material. Figure 5.11 shows an equivalent circuit of a piezoelectric transducer coupled with charge amplifier. The circuit can be considered as a piezoelectric sensor.

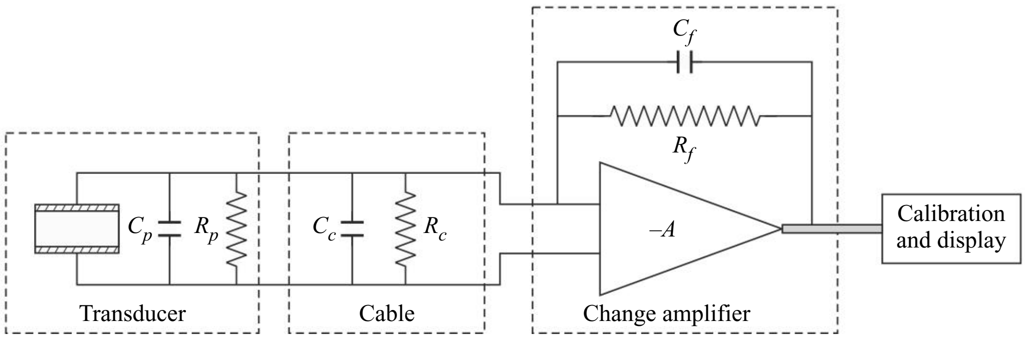


Fig. 5.11 *The piezoelectric transducer coupled with charge amplifier*

There are four sections; the transducer, the cable, the charge amplifier and the display unit. R_p, R_c and R_f are called leakage resistance, cable resistance and feedback resistance (also called time constant register), respectively. C_c and C_f are cable capacitance and feedback capacitance, respectively. A is the open loop voltage gain of the charge amplifier.

5.6.2 Piezomechanics

Piezoelectricity is a phenomenon, which implies that electricity is generated by applying force that causes mechanical deformation to any piezoelectric materials such as PZT-ceramics (Lead–Zirconia–Titanate). Reverse phenomenon is also possible, which implies that when such materials are exposed to electrical fields mechanical deformation takes place, i.e. the shape of the material changes. The effect is complementary to piezoelectricity. The amount of deformation is proportional to the applied electric field. The phenomenon is known as piezomechanics. The simplest piezomechanic component can be designed by forming a PZT layer and applying a potential difference across each side of the layer as shown in Fig. 5.12. It is a monolayered PZT component. The PZT components are also called *moving capacitors*. The monolayered PZT component can be used as an actuator. Actuation is the process of conversion of energy to mechanical form. A device that accomplishes this conversion is termed actuator.

Based on the geometrical configuration, logically there exist two types of deformations or effects. They are denoted as d_{33} and d_{31} effects. The d_{33} effect is characterized by the deformation as shown in the Fig. 5.12(a), which is the deformation in the direction of the applied field. Without loss of generality, it can be convinced that because of this d_{33} deformation there must exist a negative deformation (i.e. shrink) in another direction, called in-plane direction. The *in-plane* deformation (shrinking) is characterized by d_{31} effect. The modes of operation due to d_{33} and d_{31} effects are called thickness mode and planner mode, respectively. The deformation is governed by the physical property of the material, which is eventually reflected through its mode of operation. The above deformations correspond to two types of coefficients such as d_{33} coefficient and d_{31} coefficient, respectively. The deformation is not only proportional to the applied electric field but also to the d_{33} coefficient, if it is operating in d_{33} mode. Multilayered PZT component can also be designed to operate in the d_{33} mode. Figure 5.12(b) shows a multilayered PZT component, in which piezoelectric layers are stacked. Planner type piezoelectric microactuators has a sandwich structure as shown in Fig. 5.12(c). The voltage connections are shown. The shear mode effects are shown in Fig. 5.12(e) (detail description is presented in Section 5.12). A typical spring type piezoelectric actuating components shown in Fig. 5.12(d) are also commonly used for actuation purposes.

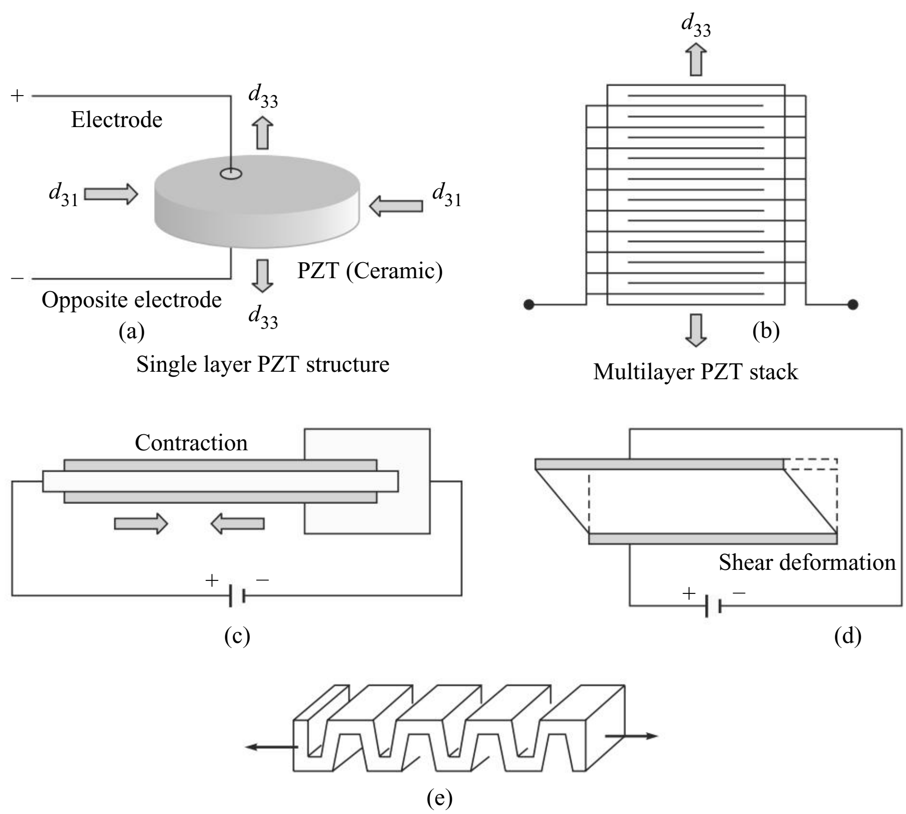


Fig. 5.12 *(a) Monolayered piezoelectric components; (b) multi-layered, (c) Planner-type sandwich; (d) Sheartype horizontally polarizes; (e) A typical spring-like piezoelectric material based microactuator construction*

Courtesy: (a) and (b) Piezomechanik GmbH; *(e)* Ounaies et al., *ICASE Research quarterly*, 8/2, 1999

Demands on next generation actuation technology include components that are smaller, lighter more energy efficient. In this respect multilayer PZT actuating components are preferred. Two important types of stacking such as *bare stack* and *ring stack* are possible. Bare stacks operate in the way the monolayered components do. The ring stacks are useful for gripping applications. The stack configuration can hold cylindrical parts and make it possible to do feed-through operation. Feed-through operations are essential in setting and handling the optical fibers and components. Stacking of several layers equivalently increases the total stroke or elongation. The stacking of such thin layers yields an actuating component, which generates large strains or deformation at relatively low applied voltages. Low voltage operation has enabled the development of state-of-the-art multilayer piezoelectric actuators. Multilayer piezoelectric ceramic actuators are capable of generating movements in the order to micrometer with tens of microseconds of response time.

Piezoelectric actuators are specified according to their stoke, pushing power, torque and weight. Practical limitations include excitation voltages, mechanical durability, interfacing capability, coupling to the control structure and control complexity, stability and sustainability to application environment.

Furthermore, the effects of dispersion at low frequencies as well as nonlinearity at high electric field strengths draws significant attention while not only designing and fabricating the piezoelectric microactuators but also the strain measuring transducers.

5.7 STRAIN MEASUREMENT

The piezoresistive property of electronic materials is frequently utilized for strain sensor fabrication. Many metal and metal alloys thin films have been studied for strain measurements. They are Au, Cu, Mn, Bi–Sb, Ni–Cu, Ni–Cr, Pd–Cr, and so on. RuO_2 based thick film can also be utilized for piezoresistive strain sensors. Polysilicon is another well-known piezoresistive material. In addition to the materials mentioned above many other materials such as berlinite ($AlPO_4$), gallium orthophosphate ($GaPO_4$), ceramics with perovskite or tungsten-bronze structures ($BaTiO_3$, $KNbO_3$, $LiNbO_3$, $LiTaO_3$, $BiFeO_3$, Na_xWO_3, $Ba_2NaNb_5O_5$, $Pb_2KNb_5O_{15}$), polymer materials such as rubber and wood fiber exhibit the effect.

5.8 PRESSURE MEASUREMENT

Piezoresistive pressure sensors are common. The simplest construction is that in which the silicon is selectively etched from a wafer in certain areas to form a diaphragm. For the measurement of absolute pressure the silicon wafer is then fused or joined with another wafer to form a vacuum-sealed cavity below the diaphragm. The diaphragm is allowed to deflect from its original position in response to the applied pressure. Piezoresistive effect is exploited as the transduction mechanism. Piezoresistive materials are implanted into the diaphragms. Because of applied force or load the deflected diaphragm causes a change in strain on the implanted piezomaterial, effectively changing the value of intrinsic resistance. The piezoresistors are implanted at the point where maximum strain or stress is exerted. U-shaped and rectangle piezoresistive cantilever arrays can also be fabricated in order to analyze not only stress/strain (because of the applied pressure), but also the presence of noise.

5.9 FLOW MEASUREMENT USING INTEGRATED PADDLE-CANTILEVER STRUCTURE

The micromachined structures can be used for the measurement of flow rate either in the differential pressure mode or dragging mode. Typical flow sensor includes three elements; a channel, cantilevers attached with a rectangular plate paddle and piezoresistive materials implanted into the cantilevers. The piezoresistive material is integrated at the root of the cantilever. The cantilever integrated with piezoresistive materials is the important part of the sensor.

The fluid or gas is passed through the channel. Because of gas flows the force is exerted on the paddle. This in turn deflects the cantilever. The deflection develops a strain resulting in a change of the resistance of the piezoresistor implanted in the cantilever at its root. A voltage output signal can be obtained by interfacing the MEMS structure with the passive electronic systems (PES) such as a Wheatstone bridge. The geometry of the paddle is usually square or rectangular in structure. Figure 5.13(a) shows the sensor has only one paddle, whereas Fig. 5.13(b) shows a sensor fabricated with two paddles. More number of paddles is essentially required for the measurement of low values of flow rate.

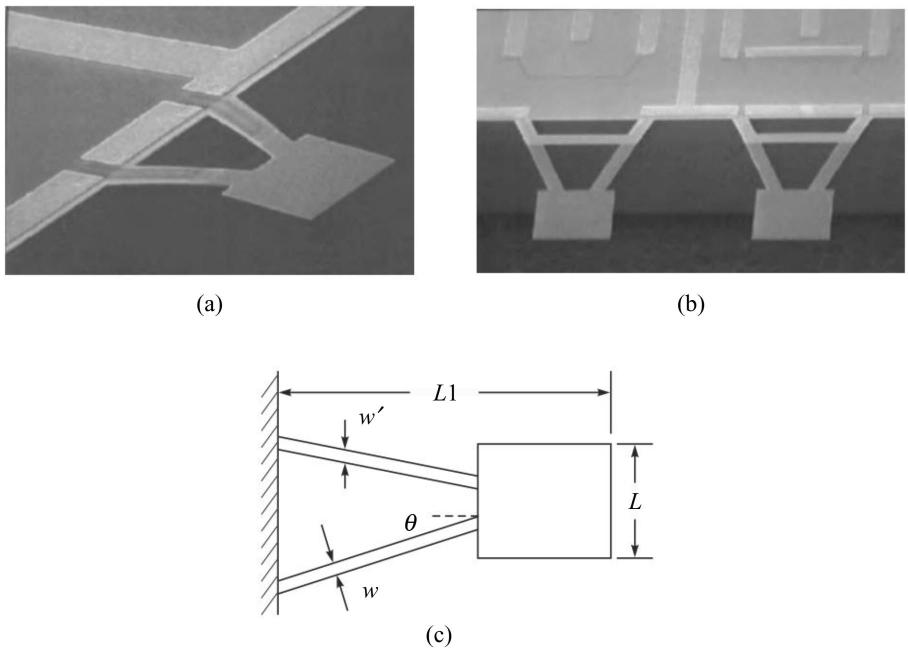


Fig. 5.13 *Cantilever and paddle-based flow sensor (a) One paddle; (b) Two paddles, (c) Schematic dimension (Length: 100–200 micrometer; thickness: 3–5 micrometer)*
Source: Su et al., *J. Micromech. Microeng.* 6, 1996

The relationship between the relative change of the piezoresistance and the deflection of the cantilever paddle at the free end is given by,

$$\frac{\Delta R}{R} = \beta \frac{3t_{pc}EL_1t}{2(L_1^3 - L^3 + 2wL^2)}\, y(0) \tag{5.10}$$

where, R and ΔR are the original resistance and change in resistance due to strain, E is the Young's modulus, t_{pc} is longitudinal piezoelectric coefficient of the material, β is a constant that signifies the positioning of the piezoresistor, $y(0)$ is the initial deflection factor, t is the thickness; L, L_1, and w are the lengths and width of the structure as shown in Fig. 5.13(c), respectively. A relatively bigger paddle square is attached to two thin and narrow cantilever arms in order to achieve a large sensitivity in flow rate measurement.

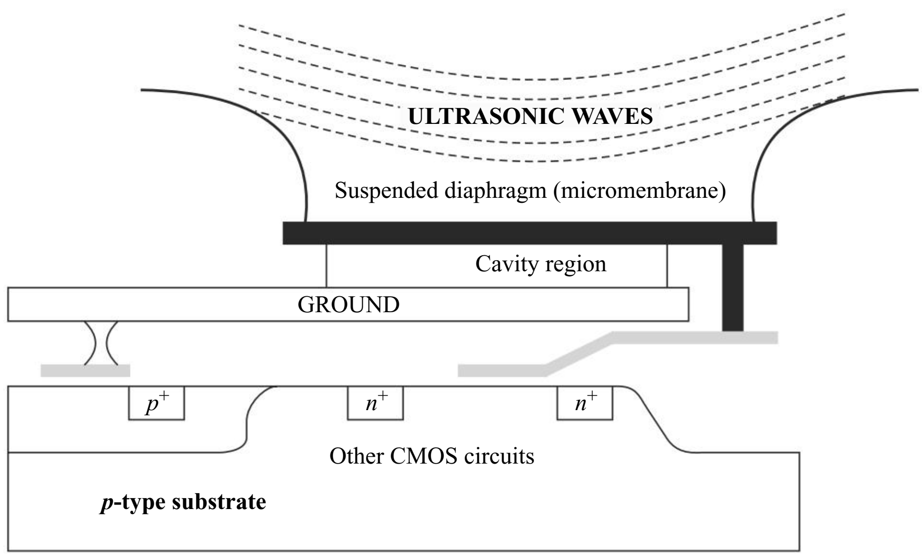
5.10 PRESSURE MEASUREMENT BY MICROPHONE

Microphone is a device that measures acoustic signal. A microphone is known as an electromechanical transducer that converts sound pressure into an electrical signal. The ordinary microphone we use in everyday life can be called macro microphone but the microphone we will be discussing here is the MEMS microphone. The macro microphones are primarily used in public address (PA) systems, on the other hand the MEMS microphones are used for acoustical measurements. There are many applications of microphones. For instance, a diver uses MEMS microphone when operating in dark murky waters. MEMS microphones are available in integrated form in which the acoustic sensing mechanism and

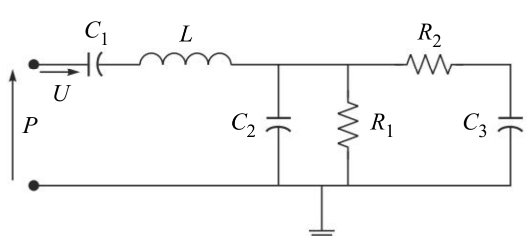
signal processing circuitry are micromachined on the same wafer. Such integrated sensors are sometimes referred to as ultrasonic sensors. The combination of signal processing circuitry and microphone itself promises reliability, integrability and compactness. The signal processing circuits are amplifier, analog to digital converter and noise filters. On-chip electronics also minimizes the effects of parasitic capacitance that might predominate if the separate external signal processing circuit would have been designed. Further, the integrated system is essentially a low voltage device and consumes less power.

5.10.1 Principle of Operation and Lumped Parameter Model

The MEMS technology based ultrasonic sensor is fabricated in a standard CMOS platform that realizes mechanical vibration of a micromembrane implemented in silicon as shown in Fig. 5.14. The

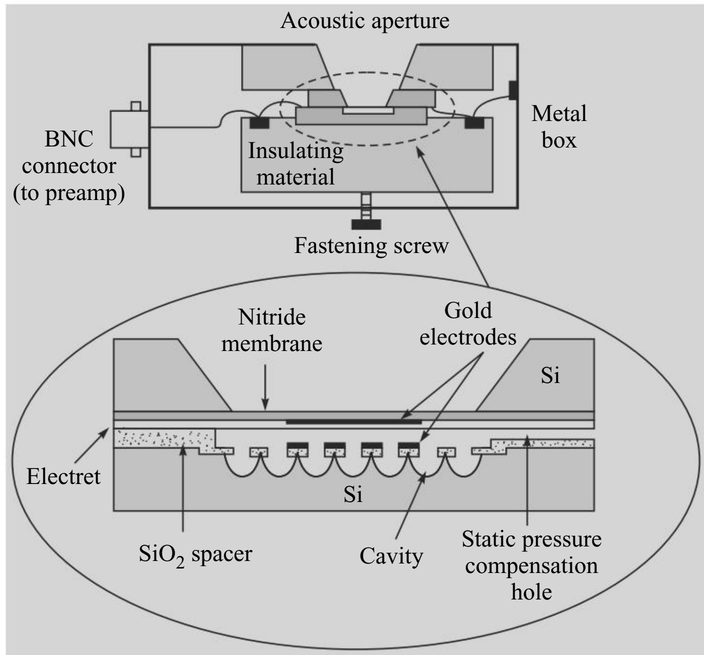


(a)



C_1 = Represents the compliance of the membrane
C_2 = Represents the compliance of the air within the cavity
C_3 = Represents the compliance of the back volume
L = Represents the mass of the membrane
R_1 = Represents the dissipative part of the compression
R_2 = Models the acoustic resistance due to viscosity

(b)



(c)

Fig. 5.14 *(a) Schematic diagram of a CMOS MEMS microphone, (b) Electrical analog to acoustic behavior of microphone system*
Source: Neumann and Gabriel—*Copyright IEEE,* Transducer 03), *(c) More detail description of cavity area of a typical CMOS MEMS microphone (Source: Hsu, Hsieh and Tai, Furutani–CIT-TTI).*

micromachined microphone employs capacitive sensing method. The transduction behavior is predicted by considering the acoustic and electrical behavior of the microphone. The acoustic behavior can be modeled by electrical lumped parameters. The sound pressure is treated as the deviation from the ambient pressure. The acousto-electric analogy is given below. Figure 5.14(b) shows the lumped parameter based model of a microphone system.

Sound pressure	:	Voltage (P)
Volume velocity	:	Current (U)
Acoustic impedance	:	Voltage/Current (P/U)
Acoustic compliances	:	Capacitance
Air mass	:	Inductance
Dissipative element	:	Resistor

The suspended micromembrane forms a capacitor with the ground plane of the wafer. The capacitance value changes in response to input, i.e. acoustical vibration or pressure. The deviation from the initial values depends on the amplitude of the incident wave. The capacitive pick-up is taken as the read out. Another way of explaining the principle of operation of such device is that it exploits changes in capacitance due to mechanical vibrations to produce voltage variations proportional to sound waves. The phase of the signal can also be measured by simply extracting the phase of the oscillations of the

capacitor. Moreover, the incident time of the incoming wave can also be known. In many MEMS microphones, dense arrays of such capacitors along with signal processing electronics are integrated in order to improve the operational capabilities. A typical multi-membrane CMOS MEMS microphone can hold an array of 64-micromachined capacitor microphones etched in silicon.

The micromembrane has charge storing capability. Fluorocarbon electrets such as Teflon material are chosen for this membrane as they have excellent charge storage characteristics and in many designs can act as a permanent charge source without the need for external biasing. The fabrication steps of microphone are as follows.

Stage-1

- Preparation of wafer
- Coating of nitride diaphragm layer of 1 μm using LPCVD
- Anisotropic back etching of wafer to form freestanding nitride diaphragm
- The front side is then coated with Cr/Au film to form the conductor membrane
- The Teflon film is deposited using spin coating
- Implantation of electrons into the Teflon film is performed at 10 keV

Stage-2

- Preparation of wafer for the back plate electrode
- Deposition of back plate electrode (thermal oxide) of thickness 3 μm
- Portion of the thermal oxide are to be etched to create windows which extend to the substrate. This forms the air gap. KOH is used as the etchant
- A 3 μm thickness of thermal oxide is again grown
- A 50 × 50 array of cavities are formed. These cavities help in reducing the air streaming resistance. Typical diameter of the cavity is 30 μm. The structure of the cavity looks like a half-dome shape with diameter and depth are 80 μm and 50 μm respectively. The cavities are formed through etching
- Lastly Cr/Au electrode is deposited using a physical mask

5.10.2 Piezoelectric Membrane

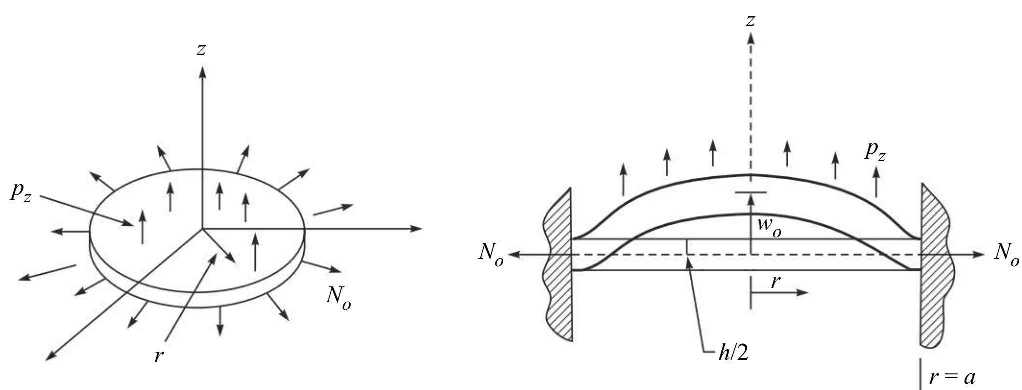
There are certain MEMS microphones in which piezoresistive materials are used as the membrane. In a piezoresistive sensor the resistance is modulated. A piezoresistive membrane transforms acoustic energy into electrical energy by developing stress or strain on the top surface of the membrane, which is deformed by the acoustic input pressure, into a corresponding change in resistance. For a piezoresistive microphone, it is essential that the stress must be linearly related to the input acoustic pressure. Other notable parameters are mechanical sensitivity and bandwidth. Quantitatively, the modulated resistance is linearly related to the product of the applied stress on a piezoresistor and the piezoresistive coefficients and is given by,

$$\frac{\Delta R}{R} = \sigma_l l_{pc} + \sigma_t t_{pc} \tag{5.11}$$

where, the subscripts l and t indicate the longitudinal and transverse directions with respect to resistor orientation, respectively. The linearity, sensitivity and bandwidth of the membrane are directly related to the structural design of the membrane.

Let us consider a circular piezoelectric membrane fixed at periphery, as shown in Fig. 5.15, has undergone a uniform in-plane stress. Let us define the following parameters.

a = radius of the membrane
h = thickness
p_z = transverse pressure
$\phi(r)$ = slope of the membrane because of deformation
$w(r)$ = transverse deflection
E = modulus of elasticity
ν = Poisson's ratio
k = tension parameter
$I_o(\,)$ & $I_1(\,)$ = Bessels functions of first kind
$\sigma_r(r)$ = radial stress
$\sigma_t(r)$ = tangential stress



(a) (b)

Fig. 5.15 *MEMS microphone diaphragm*
Courtesy: American Institute of Aeronautics and Astronautics Inc.

Now, Eq. 5.12 to Eq. 5.15 provide the expressions for the slope, the transverse deflection, radial stress and the tangential stress on the top of the membrane, respectively.

$$r^2\frac{d^2\Phi}{dr^2}+r\frac{d\Phi}{dr}-\left\{1+k^2\left(\frac{r}{a}\right)^2\right\}\Phi = 6(1-\nu^2)\frac{p_z r^3}{Eh^3} \tag{5.12}$$

$$w(r) = \frac{12(1-\nu^2)\,p_z a^4}{k^2 Eh^3}\left\{\frac{I_0\left(\frac{kr}{a}\right)-I_0(k)}{2kI_1(k)}+\frac{a^2-r^2}{4a^2}\right\} \tag{5.13}$$

$$a_r(r) = \frac{3p_z a^2}{h^2}\left\{\frac{\nu+1}{k^2}-\frac{I_0\left(\frac{kr}{a}\right)}{kI_1(k)}-\frac{a(\nu-1)I_1\left(\frac{kr}{a}\right)}{rk^2 I_1(k)}\right\} \tag{5.14}$$

$$a_t(r) = \frac{3p_z a^2}{h^2}\left\{\frac{\nu+1}{k^2} - \frac{\nu I_0\left(\frac{kr}{a}\right)}{kI_1(k)} - \frac{a(1-\nu)I_1\left(\frac{kr}{a}\right)}{rk^2 I_1(k)}\right\} \tag{5.15}$$

$$k = \frac{a}{h}\sqrt{\frac{12(1-\nu^2)\sigma_0}{E}} \tag{5.16}$$

The equations can be considered as model equations and, therefore, can be used for analyzing the behavior of the system during simulation study, i.e. prior to the design of the real micromachined MEMS microphone.

5.11 MEMS GYROSCOPES

The design of MEMS rate sensors using silicon-processing techniques took sufficient time to emerge. Rate sensors are those sensors by using which primarily the angular movement of a body can be measured. Angular rate is simply a rotational speed that implies how quickly an object turns. Angular rate can be specified in various units such as rotations per minute (RPM), degrees per second, radians per second (1 RPM = 6 deg/s; 1 rad/s = 57.3 deg/s). The suitable structural system for the measurement of angular movement is a gyroscope.

Gyroscopes (or simply gyros) are of two types: optical and mechanical. In optical gyros two laser beams are rotating opposite directions in a closed loop. When the beams are combined the rotation rate and direction can be calculated from the interference fringes. Since this chapter deals with only mechanical sensors, optical gyros will not be described. Mechanical gyroscope operates based on conservation of momentum and measures the changes in linear or angular momentum. Mechanical gyroscope utilises continuous angular movement and is therefore referred to as momentum-flywheel gyroscope. There are two basic classes of rotation-sensing gyros, such as rate gyros and rate integrating gyros. In rate gyros the output is relative to the angular speed whereas in rate integrating gyros the produced read out is the actual turn angle.

Conventional macro gyroscopes shown in Fig. 5.16 are very expensive and large in size, and consist of mainly four components such as a base (not shown), two rings and a massive rotor. The rings are called *gimbals*. The rotor is mounted inside the rings as shown.

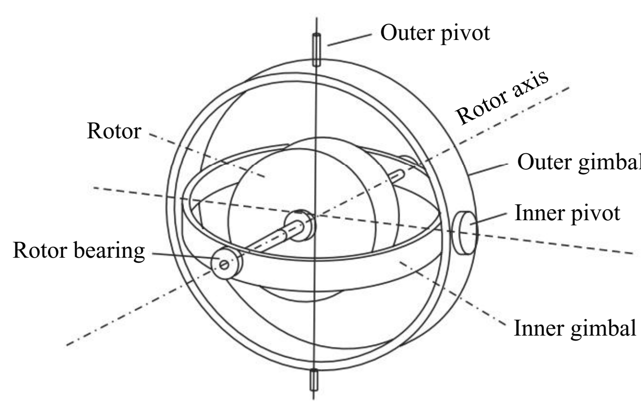


Fig. 5.16 *A conventional gyroscope structure*

If a gyroscope is pushed (tilted) at the top outer pivot, keeping the lower outer pivot balanced, while the massive rotor is moving around the axis, the gimbals will try to reorient so as to keep the spin axis of the rotor in the same direction. The reorientation hence the displacement of the gimbals can be the measure of angular speed.

The constructional design of a typical MEMS gyroscope is very different from the conventional type, however, the exploitation of the scientific principle associated with the conventional gyroscope is

equally applied to the MEMS-structured gyroscope. The conventional gyroscopes have astonishing sensitivity and can be used for space applications such as the study and measurement of the effect of the earth rotation, monitoring at micro-gravity environment, and so on. No such highly sensitive MEMS gyroscopes have yet been fabricated that are capable of measuring the desired parameters for space applications. MEMS gyroscopes are a relatively new research field, and there is still much work to be done. However, a wide range of MEMS gyroscopes are already in the technology marketplace which have an abundance of applications in the following fields.

- Control of vehicle dynamics (Automotive safety systems)
- Vibration monitoring for machines and engines
- Navigation and guidance systems
- Rollover detection
- Crash sensors
- Flight control and test instruments (inertial navigation systems for airplanes) and missiles
- Cockpit display stabilization in aircrafts
- Navigation of mobile robots in tunnels or piping

Since the demand for MEMS gyroscope is skyrocketing, a fundamental study is of paramount importance. For instance, MEMS gyroscopes or microgyroscopes are now not just for navigating airplanes anymore but can help tour buses find their way through the streets and may someday be standard equipment on millions of cars as reported by Analog Devices Inc. The principle, concepts and structural design are given below.

5.11.1 Principle

As already mentioned, the gyroscope is a sensor used to detect angular motion. The microgyroscope operates based on the principle of detecting the induced *Coriolis acceleration*. The Coriolis acceleration can be best understood from the following example. Consider Fig. 5.17, which is a flat platform spinning in clockwise direction in an axis that is perpendicular to the plane of the paper (z-axis). Assume a vibrating object is placed on the spinning platform and the object is trying to maintain its position relative to the paper. Since the object is vibrating perpendicularly, at a given speed of the platform the object would require higher speed when it is more toward the periphery, compared to the center. Higher speed implies higher acceleration. This effect is called Coriolis acceleration.

Now let us fix an MEMS gyroscope on this spinning platform, as is usually mounted in most of the locomotive vehicles. The gyro inside a vehicle measures how quickly the vehicle turns. The gyro can sense if the car is spinning out of control and therefore can take remedial actions in terms of activating the suspension and braking systems to bring it back under control. Mathematically, the Coriolis acceleration is given by,

$$a = 2\Omega \times \dot{x} \tag{5.17}$$

where, Ω is the angular velocity and $\dot{x}$ is the linear velocity (y-direction). But in general, the displacement, $S = xi + yj + zk$, and the velocity, $V = ds/dt = (dx/dt)i + (dy/dt)j + (dz/dt)k = xi + yj + zk$. Thus the velocity along y direction is denoted by y or v or dy/dt.

The angular rate is measured by sensing the Coriolis acceleration. There are various techniques in sensing the Coriolis acceleration. If the linear velocity is expressed as $V \sin\phi$, then from Eq. 5.17 the corresponding Coriolis force can be written as,

$$F_c = 2m\, \Omega V \sin \phi \tag{5.18}$$

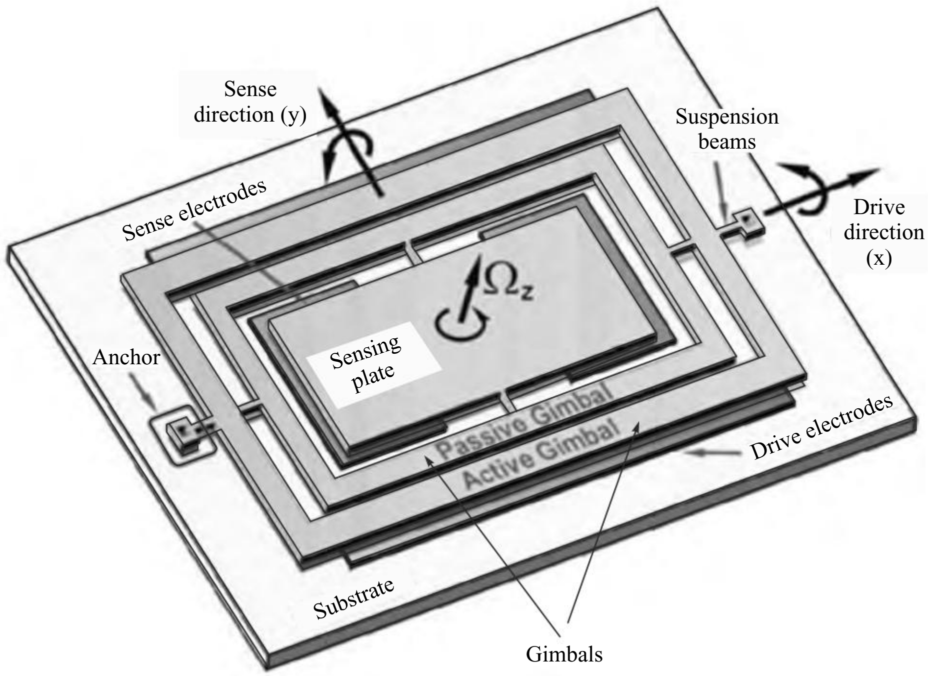


Fig. 5.17 *Principle of operation*

The Coriolis acceleration/force is measured either by optical, capacitive or piezoresistive principle. Note that Coriolis force is orthogonal to both linear motion (vibration direction) and the axis about which the rotational motion is to be sensed. A typical MEMS design of microgyroscope is shown in Fig. 5.18. Figure 5.19 shows the schematic circuit connections of a gyroscope with essential components.

5.11.2 Compensation in Gyroscope

The major drawback in designing the MEMS gyroscope is the restrictive tolerance in microfabrication. Fabrication tolerance concerns with the threshold stress. The materials used for fabrication undergo catastrophic failure beyond their elastic limits. Because of the restrictive fabrication tolerance, structural imperfection occurs. The imperfection introduces scale factor and drift errors thereby reducing the overall performance of the gyroscope. The two perturbations to the gyroscope are called regular perturbation and non-regular perturbation. Regular perturbation arises from structural imperfections and also from dissipation. On the other hand, the non-regular perturbation appears because of intrusion of three noises, namely Johnson noise, flicker noise and shot noise. Johnson noise is defined as the thermally induced electrical noise in the resistive type elements. Flicker noise originates from the contamination in the processing of materials. It can be reduced by cleanliness practices. However, it still persists. Flicker noise occurs at low frequencies and it is temperature and frequency dependent. The noise caused by random fluctuations in the motion of charge carriers in a material is called shot noise.

The imperfection can be corrected through feedback control mechanism. If the errors are large as compared to the measured Coriolis force, the feedback mechanism based correction would not solve the problem. A reliable solution is to consider the compensation method through self-calibration and signal processing technique. The compensation architecture can also embed both feedback mechanism

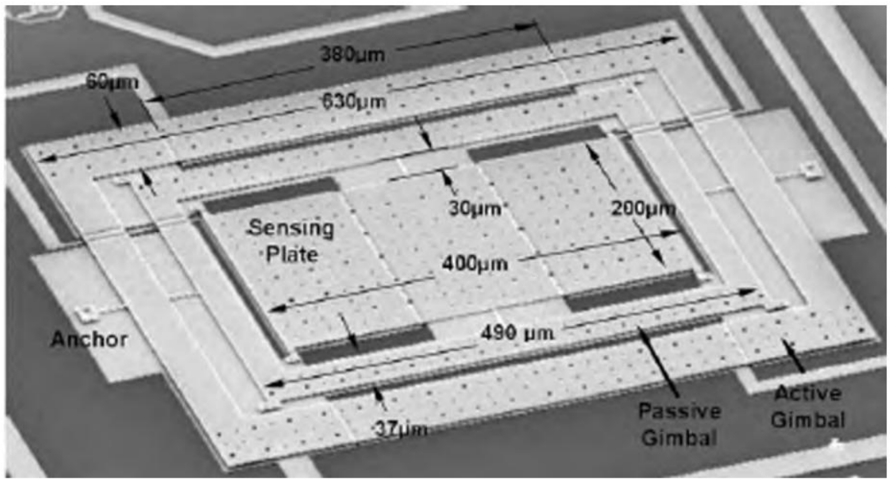


Fig. 5.18 *A practical microgyroscope*
Source: Acar and Shkel, *J. of Micromech. Microeng,* 14, 2004 – Copyright Inst. of Physics Publishing

as well as the self-calibration and signal processing technique. Such an architecture is known as dual stage architecture (DSA). A DSA is illustrated in Fig. 5.20. The MEMS gyroscope with built-in facilities for compensation of unwanted perturbations is defined as smart device. The smart device is a system-on-a-chip (SoC) device.

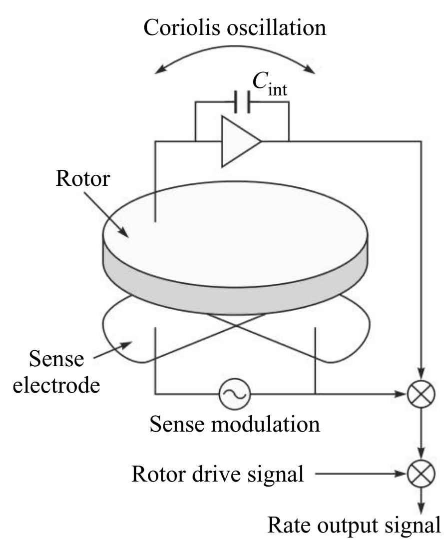


Fig. 5.19 *Circuit connections gyroscope with essential components*

The DSA has two sections one handles the small regular fluctuations and other compensates the large regular perturbation. The section that deals with large regular perturbation utilizes feedforward mechanism. This section has a sub-section that incorporates self-test algorithm for self-calibration. The self-calibrated signal is then applied for the correction of the imperfections. Note that the self-test algorithm requires the feedback signal, which is obtained from the feedback sensory elements. The section that deals with small fluctuations also requires feedback signals that are obtained from the built-in electronic sensory devices, as shown. Position and velocity feedback control is essentially used. The gyroscope is a sensor and therefore it should have output terminals to be connected to the display unit in order to display the measured quantity, i.e. angular information (tilt). The role of signal conditioning circuitry is to deliver the appropriate level of measured signal to the display unit.

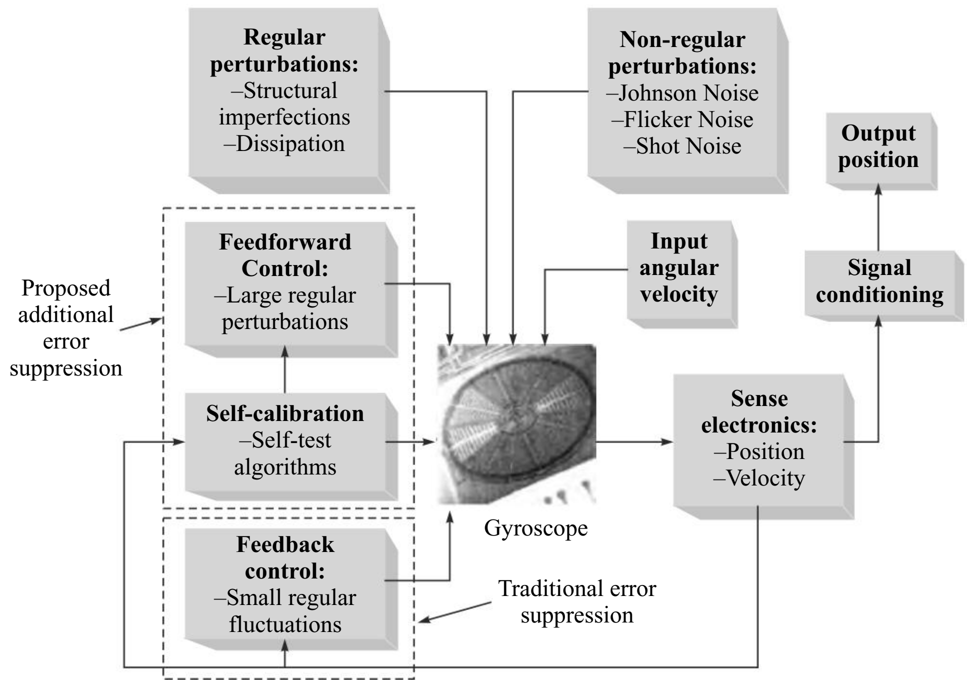


Fig. 5.20 *Smart MEMS gyroscope would include self-calibration capabilities built on-chip for detection and compensation of deterministic perturbations*
Source: Painter et al., *IEEE Sensor J.* 3/5, 2003.

5.12 SHEARMODE PIEZOACTUATOR

Figure 5.12(d) illustrates the schematic connection and its deformation in a shear mode piezoelectric actuator. In this case the piezoelectric component has to be horizontally polarized. If a voltage is applied to a horizontally polarized material then shearing force is developed that causes the element to shear. Because of shearing effect the element will be subjected to displacement, which is given by,

$$D = nd_{15}V \tag{5.19}$$

where, D is the length of displacement, n is the number of layers (if any), V is the applied voltage and d_{15} is called shear mode piezoelectric constant. Point to be noted is that the displacement is independent of the dimension of the piezoelectric materials. Therefore, MEMS conformant very thin layers and small structures can be designed for wide range of applications. Fujitsu has developed a typical shearmode piezoelectric microactuator that can be used in hard disk drive (HDD) of a computer. In traditional non-MEMS actuator system the servo bandwidth is limited by the mechanical resonances of the carriage, coil, and ball bearing pivot. Some types of microactuators have been proposed as possible ways to attain a wider servo bandwidth. The microactuator design for HDD applications can be of three types based on the way the actuator drives the component of question: driving a head suspension assembly, driving a slider, and driving a head element. Head suspension driving microactuators are easy to manufacture, they are expected to be used in HDDs in the near future in spite of their poor mechanical characteristics

compared with the slider and head element driving types. The detail structural design is shown in Fig. 5.21 and its specification is shown in Table 5.2. The complete design consists of the following parts.

- Carriage arm
- Stator plate
- Electrodes
- Single layer piezoelectric element
- Head mounting block
- Head suspension

The microactuator has a high resonant frequency of 9 kHz, with peak gain at resonance is 20 dB and with high shock resistance. These parameters are considered as the required mechanical characteristics of a good actuator.

Table 5.2 *Specification*

Mass (Head suspension)	62 mg
Stroke	0.5 µm(± 3 0V)
Resonant frequency	9 kHz
Capacitance	650 pF
Shock resistance	>950 G, 1 ms half-sine
Dimension	
Length	= 2.2 mm
Width	= 1.3 mm
Thickness	= 0.15 mm
Mass = 62 mg	
d_{15} = 8.45E–10 m/V	
Range = ± 25 nm with ± 30 Volt	
Mechanical advantage = 20 times	
Displacement at the head ≈ 1 μm_{p-p}	

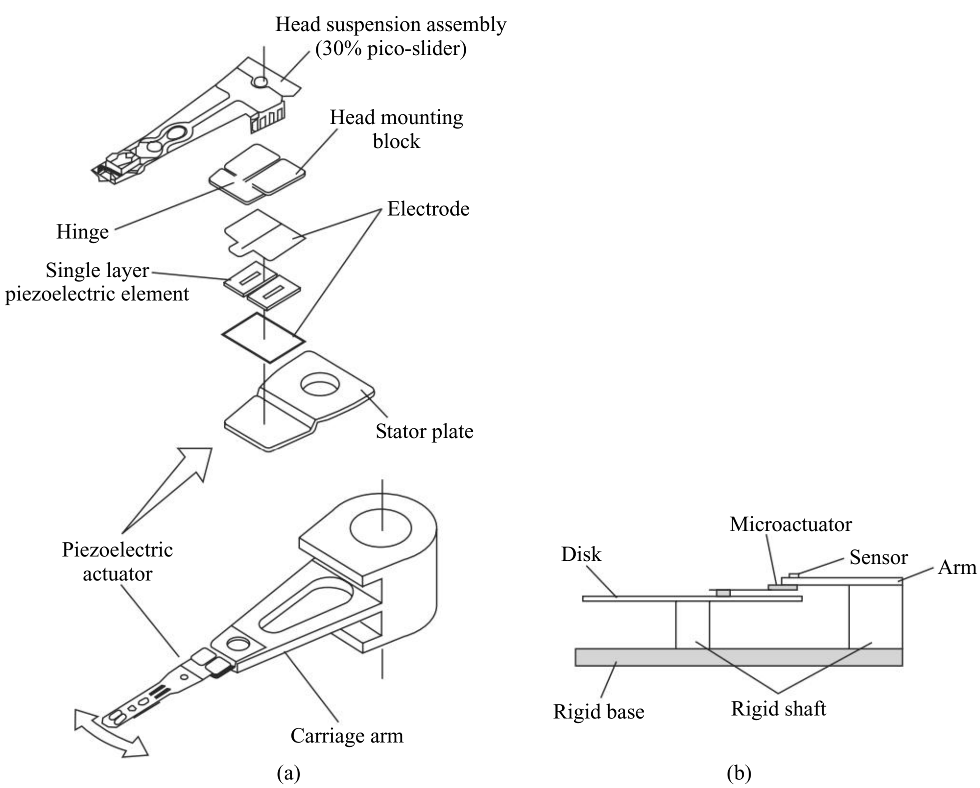


Fig. 5.21 *Piezoelectric microactuator for precise positioning of hard disk drive; (a) Various parts; (b) Integrated into the hard disk drive*
Courtesy: Fujitsu, *Source:* Koganezawa and Hara, *Fujitsu Sci. Tech. J*, 37/2, 2001

5.13 GRIPPING PIEZOACTUATOR

One of the MEMS actuation applications is found in microassembly, where the components or parts of the order of micro or nanometer are handled effectively. The parts could be a microoptical, a microelectronical or a micromechanical component. In order to handle these components several types of microgrippers have already been designed. Fraunhofer-Institute of Reliability and Microintegration and Technical University of Ilmenau has designed a microgripper, which is based on the principle of piezoelectric actuation (Fig. 5.22). Microgripper compliant mechanisms are very complex with regards to assembly of various parts. Description of various units of a typical gripper is as follows (Refer to Fig. 5.22).

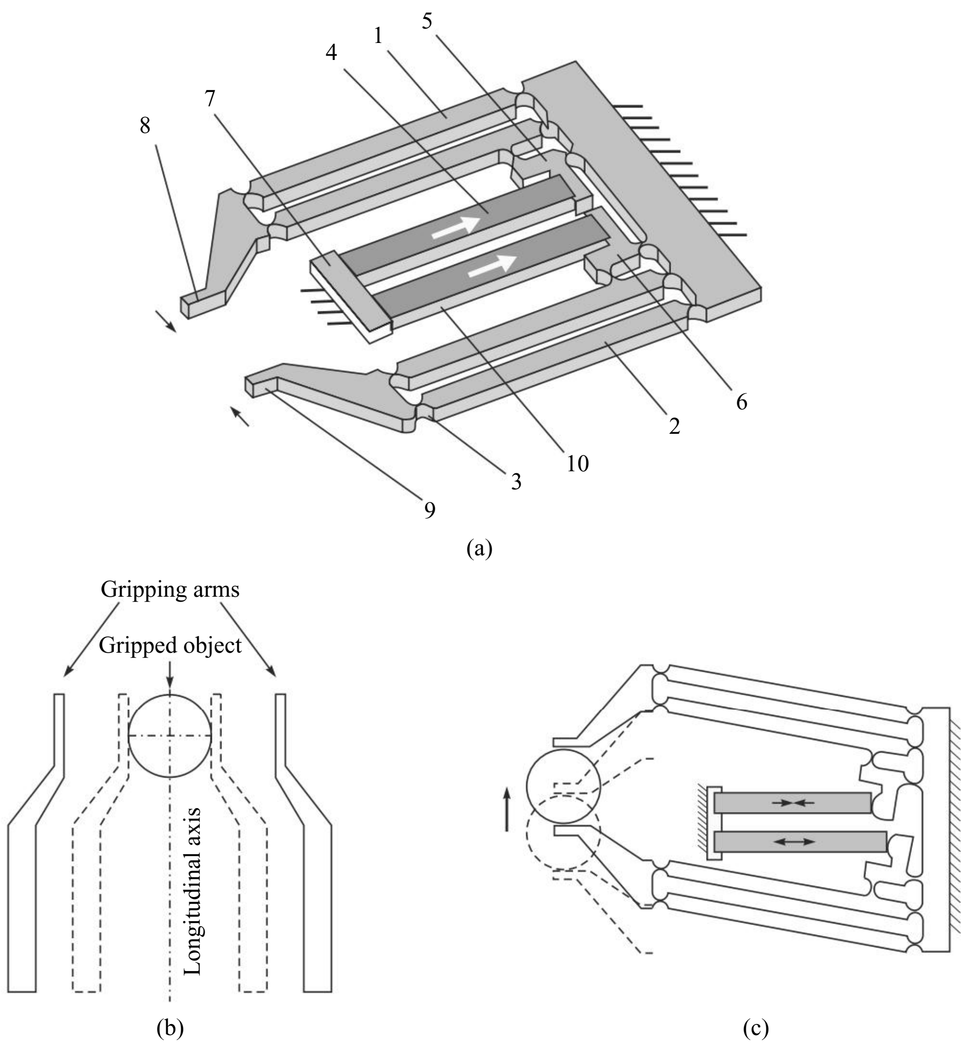


Fig. 5.22 *(a) A piezoelectric principle based microgripper; (b) Parallel movement of the arms; (c) Application of equal but opposite voltage*
Courtesy: MESSE Bremen GmbH, *Source:* Keoschkerjan and Wurmus, Actuator 2002

1. Right hand parallelogram mechanism
2. Left hand parallelogram mechanism
3. Flexure hinges
4. *Right-side piezoelectric material*
5. Right hand lever mechanism
6. Left hand lever mechanism
7. Mounting base of both the piezoelectric materials
8. Right hand gripping element
9. Left-side gripping element
10. *Left side piezoelectric material*

The principle of operation of this microgripper is as follows. The two pieces of piezoelectric materials (right-side and left-side listed at 4 and 10 respectively) causes the movement of the gripper in parallel. When voltage is applied on the piezoelectric material deformation takes place. The deformation is transmitted along the direction as shown by arrows. The deformation is transmitted to the grippers through the lever and hinges. A lever is a rigid element that is used with an appropriate fulcrum or pivot point to multiply the mechanical force that can be applied to another unit. Hinge is a point in the structure at which a member can rotate freely. The gripper basically does two operations: gripping operation and release operation. Applying equal amount of voltage to both the piezoelectric materials gripping operation is carried out. Resetting the applied voltage performs the release operation. The amount of movement is proportional to the applied voltage. Depending upon the dimension such as diameter, size and length of the components to be gripped, the value of voltage is determined. The applied voltage is called the controlling signal. By applying equal but opposite voltage to the piezoelectric material, another form of parallel deformation is achieved as shown in Fig. 5.22(c). In either case the grippers move parallel to each other. The parallelogram, lever mechanisms in conjunction with the hinges are called amplifier. The structure must be hinged in such a way that the piezoactuator deformation should get amplified at least 50–100 times.

5.14 INCHWORM TECHNOLOGY

In 1964, Sibitz and Steele were among the first to patent (Patent No. 3,138,749) an inchworm-type actuator as a low-force, high precision positioning device capable of controlling a movement that is an order of micro-inch range. Later on many educational research centers and MEMS manufacturing companies have considered the inchworm-type actuation mechanism as a technology and since then many patents have been awarded. Inchworm-motors are recent developments, which are based on inchworm technology. Inchworm technology makes use of piezoelectric material. As it stands, a machine that converts electrical energy into mechanical energy is known as a motor. Principally, an inchworm motor converts electrical energy into mechanical energy by the use of piezoelectric materials. Without loss of generality, the inchworm motors can be called inchworm actuators and vice versa. Linear motion characterized by steps as small as nanometers and an overall range of travel of hundreds of microns is achieved from the inchworm actuator.

5.14.1 Principle of Operation

The inchworm technology is based on a sequence of clamping-releasing mechanism by the use of a series of piezoelectric materials subjected to controlled electrical signal (voltage). (Refer to Fig. 5.23.) For the purpose of understanding, a simple schematic diagram has been drawn, however, the actual MEMS design is extremely complex. There are three sets of piezoelectric materials pm_1, pm_2 and pm_3. pm_1 and pm_2 are called clampers. pm_3 is called mover. When electric potential is applied separately they change their shapes; deform (elongates) in a direction as shown in the figure. The design, positioning and orientation is such that the clamper pm_1 and pm_2 always elongate in a vertical direction and pm_3

elongates in the horizontal direction. The electric potentials are applied to these materials in a proper sequence, as will be depicted below. At this point one should take a note that the dark color implies that the piezoelectric material is under deformation mode caused by the applied electric potential, and clear implies that no voltage has been applied.

Initially, with zero applied voltage the shaft of the motor, which is usually set aside to the pm_1, pm_2 and pm_3 is at rest. Now, a calculated amount of DC voltage is applied to pm_1 in order to force it to act as a clamper, and in fact it clamps the shaft. Note that because of its deformation in the vertical direction, the clamping operation is accomplished. The required voltage has to be calculated beforehand in order to achieve perfect clamping. Overclamping and underclamping may introduce stress penalty that is unwanted. In the next sequence, the middle, i.e. pm_3 is activated. Recall that it elongates in the

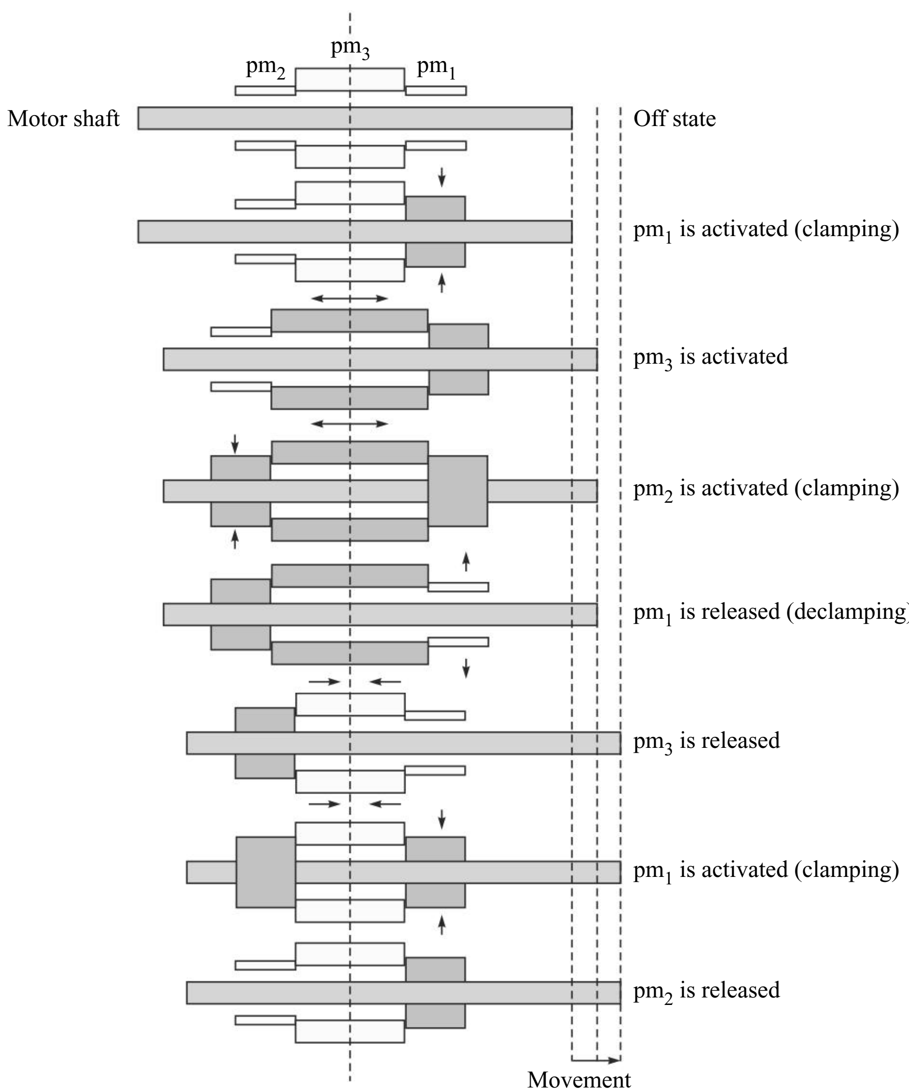


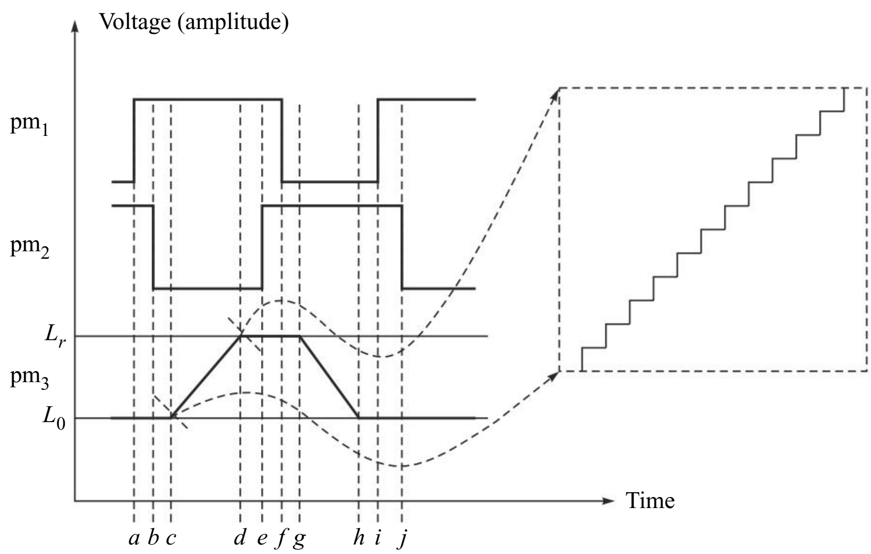
Fig. 5.23 *Inchworm-principle based linear movement*

horizontal direction. Because of its deformation, the motor shaft moves toward the right. The sequence of clamping-release activation is given below (Fig. 5.23).

1. No activation (OFF state)
2. pm_1 is activated (ON state)
3. While pm_1 is ON, pm_3 is activated
4. **A movement is created**
5. While pm_1 and pm_3 are both ON, pm_2 is activated
6. While pm_2 and pm_3 are both ON, pm_1 is deactivated
7. While pm_2 is ON, pm_3 is deactivated
8. While pm_2 is ON, pm_1 is activated
9. While pm_1 is ON, pm_2 is deactivated
10. **Another movement is created**

5.14.2 Controlling Signal

Providing appropriate voltage signal to the pm_3, the position of the shaft can be controlled. The voltage applied to pm_3 is usually an incremental step voltage and is called driving voltage. The amount of movement and the direction of movement are both controllable through controller. Figure 5.24 illustrates the timing diagram as far as clamping-releasing operations are concerned. The figure is self-explanatory, if you follow the sequence of activation mentioned above.



At time, a = pm_1 is clamped (+ve voltage)
At time, b = pm_2 is not clamped (–ve voltage)
Time, c to d = pm_3 is step-incremented from the level L_0 to L_r, depending upon the positioning requirement (Elongation)
At time, e = pm_2 is clamped (+ve voltage)
At time, f = pm_1 is unclamped (–ve voltage)
Time, g to f = pm_3 is step-decremented (Contraction)
Time, a to f = clamping timing of pm_1
Time, e to j = clamping timing of pm_2

Fig. 5.24 *Timing diagram of a typical inchworm microactuator*

The clamping durations of both pm_1 and pm_2 is longer than the elongation and contraction duration of the pm_3. During contraction the shaft also does a movement. A single step in the driving signal causes a movement. The required linear movement depends on the amplitude of L_r. The shaft can also be moved in the opposite direction, i.e. leftward, simply by reversing the process, i.e. by activating pm_2 in first place and following a sequence similar to the previous one. Increasing the frequency of clamping and unclamping signals the speed of movement can be increased. The important aspect of driving mechanism is the accurate synchronization, since the movement is mainly a step-by-step sequential driving procedure. The inchworm motion is thus achieved by a combination of piezoelectric driving and electrostatic clamping. Inchworm-principle based actuator have large output forces.

A typical inchworm technology based position actuator can make a movement of 100 microns. The power dissipated by the inchworm actuator is given by,

$$P = CV^2f \tag{5.20}$$

where, C is the total capacitance of the actuator, V is the applied voltage, and f is the frequency of operation. The inchworm technology can be used for the design of gait for microrobots. An inchworm-like robot is often called a mobile microrobot that reproduce the locomotion pattern of a natural inchworm. They consist of integrated actuating modules that can deform in the direction of travel. The locomotion is based on a series of cyclic actuator actions. There are several applications of inchworm microrobots. Very specific and typical application includes inspection and material delivery tasks in narrow and highly constrained environments.

Primarily, the MEMS inchworm devices are used for micropositioning applications. There are plenty of precise positioning applications where inchworm motors are preferred because of many added advantages. For example, positioning of microelectrodes for electrophysiology recordings require micrometer-scale control of high-velocity, high-acceleration incremental steps in forward and backward direction. Moreover, the positioning must be achieved without vibration so that cells can be under protection. Other applications include recording (recording requires micro-stage positioning of head), positioning of optical components, positioning of microsurgical equipments, cell penetration, microinjection, suppression of noise and vibration in scientific instruments and friction measurement.

5.15 SUMMARY

Mechanical MEMS emphasizes two classes of devices; mechanical structure based device and piezoelectric material based device. Various geometric structural configurations are cantilever, beam, plate, diaphragm and hollow chamber. Piezoelectric material based mechanical sensors and actuators exploit the effect of piezoelectricity. Broadly, the mechanical MEMS utilize the following methods and principles.

- Cantilever beam as sensing element
- Use of plates/diaphragm structure for capacitive sensing
- Microphones as sound sensor
- Exploitation of Coriolis acceleration in gyroscopes for angular rate measurement
- Principle of piezoelectricity

MEMS mechanical sensors are very popular because of easy integration procedure in the micromachining process. The basic challenge that is encountered in designing, however, is the implementation of signal processing circuitry. Further, it inherits many disadvantages that are given below

- The overall silicon area is generally larger.
- Larger volume with respect to packages.
- Multi chip modules require additional integration steps.
- Existence of larger signals from the sensor output.
- Stray capacitance of the interconnections.

Points to Remember

- Sensors and actuators play a significant role in almost all kinds of instrumentation, process control, factory automation and machine control applications.
- Micromechatronic systems are mechatronic systems but their dimensions are measured in microscale levels.
- Sensing is performed by using transducers.
- Transducers are mainly used to quantify the physical, electrical, fluidic and mechanical *variables* such as temperature, pressure, magnetic field, voltage, flow, vibration, radiation, and so on.
- The term transducer and sensor have been used synonymously although the concepts are little different in the macro level system design. Transducers are the physical element, which is a part of a sensor.
- Actuation is simply the process of conversion of one form of energy to mechanical form. The process is known as the principle of energy transduction. A device that accomplishes this conversion is termed actuator.
- The important microactuators are Mechanical actuators, Thermal actuators, Electrostatic actuators and Magnetic actuators.
- Mechanical microactuators are broadly categorized into two groups, namely (i) Mechanical structure based microactuator and (ii) Active material based microactuators.
- Materials, which undergo some sort of transformations through physical interactions, are referred to as *active materials*.
- Some of the important active materials are *piezoelectric* material and *magnetostrictive material*.
- Piezoelectricity concerns mechanical properties such as stress and strain.
- The mechanical MEMS utilize the following methods and principles: Beam and cantilever methods; Capacitive method and Piezoelectric principle.
- Various structures and features of MEMS plates are square, rectangular, solid, porous, single-layered and multi-layered.
- At macro level, diaphragms are thin circular plates broadly used for the measurement of both low and high values of pressure, force or load. The principle is based on deflection.
- CMOS-based ultrasonic MEMS diaphragm is used for fault detection: by sensing the echo of an ultrasonic burst in a structure. In such application acoustic impedance of the diaphragm is exploited.
- Basically, the geometrical dimensions as well as the property of the dielectric material of a parallel plate type capacitor are exploited to quantify many types of measurands. The principle of capacitive transducer is based on the change of *distance*, *area* and/or the *permittivity*.
- Piezoelectric materials are permanently polarized crystalline materials, if their dimensions are changed as a result of mechanical force, electric charges proportional to the imposed force are accumulated on the surface upon which the force is imposed. The phenomenon is known as the piezoelectric effect or *piezoelectricity*, which was originally studied by Pierre and Jacques Curie in 1880.
- Piezoelectric transducers are cut from the crystal slabs or plates in a specific orientation with respect to the crystal axes depending on the application requirement.
- The amplifier used in the piezoelectric sensor is referred to as *charge amplifier* due to the reasons that the accumulation of the charge effectively appears as a difference in the potential across the material.

- Piezoelectricity is a phenomenon which implies that electricity is generated by applying force that causes mechanical deformation to any piezoelectric materials such as PZT-ceramics.
- Based on the geometrical configuration of the piezoelectric material, logically there exist two types of deformations or effects. They are denoted as d_{33} and d_{31} effects.
- Piezoelectric actuators are specified according to their stoke, pushing power, torque and weight.
- The micromachined structures can be used for the measurement of flow rate either in the differential pressure mode or dragging mode. Typical flow sensor includes three elements; a channel, cantilevers attached with a rectangular plate paddle and piezoresistive materials implanted into the cantilevers.
- Microphone is a device that measures acoustic signal. A microphone is known as an electromechanical transducer that converts sound pressure into an electrical signal.
- The MEMS technology based ultrasonic sensor is fabricated in a standard CMOS platform that realizes mechanical vibration of a micromembrane implemented in silicon. The capacitive pick-up is taken as the read out.
- Rate sensors are those sensors by using which primarily the angular movement of a body can be measured. Angular rate is simply a rotational speed that implies how quickly an object turns.
- Gyroscopes (or simply gyros) are of two types, optical and mechanical.
- Mechanical gyroscope operates based on conservation of momentum and measures the changes in linear or angular momentum.
- The microgyroscope operates based on the principle of detecting the induced *Coriolis acceleration*.
- The major drawback in designing the MEMS gyroscope is the restrictive tolerance in microfabrication.
- The two perturbations to the gyroscope are called regular perturbation and non-regular perturbation. MEMS gyroscopes with built-in facilities for compensation of unwanted perturbations are defined as smart devices.
- The imperfection that usually occurs in MEMS gyroscope can be corrected through feedback control mechanism.
- Microgripper compliant mechanisms are very complex with regards to assembly of various parts. The gripper can be used to generate movement.
- The inchworm technology is based on a detrimental sequence of clamping-releasing mechanism by the use of a series of piezoelectric materials subjected to controlled electrical signal (voltage).
- There are three sets of piezoelectric materials pm_1, pm_2 and pm_3. pm_1 and pm_2 are called clampers. pm_3 is called mover.
- A typical inchworm technology based position actuator can make a movement of 100 microns.
- The power dissipated by the inchworm actuator is given by $P = CV^2f$.

Exercises

1. Why are sensors designed?
2. Define an actuator. With a simple diagram illustrate the connections between the sensors, actuator, microprocessor (or microcontroller) and the target machine platform.
3. List out the physical parameters that can be sensed by using sensory devices.
4. What do you mean by 'equivalence' in case of a transducer element? How does a sensor device differ from a transducer element?
5. Write down the principal energy transduction methods that are employed for achieving actuation. Also briefly explain their principle of operation. Write down some of the important applications of microactuators.

6. Suggest the principle of sensing and actuation with respect to mechanical MEMS.
7. Write short notes on the following mechanical MEMS elements.
   (a) Beam and cantilever
   (b) Microplate
   (c) Diaphragm
8. Discuss the theory and principle of operation of a typical parallel plate type capacitive sensor.
9. What do you mean by piezoelectricity? Name some of the piezoelectric materials. Define piezoelectric constant of a piezoelectric material. Why is the amplifier used in piezoelectric sensor called charge amplifier? Draw the circuit diagram of a typical charge amplifier and define the components.
10. What do you mean by piezomechanics? How does it differ from piezoelectricity? Explain d_{33} and d_{31} effects in a piezoelectric material.
11. Explain the principle of operation of a paddle-cantilever structure based flow sensor, integrated with piezoresistive materials.
12. Discuss the principle of operation and the constructional features of a typical MEMS microphone that can be used for the measurement of acoustic signal (sound signal). Assume that the diaphragm of the microphone is made up of piezoelectric materials. Now write down the expressions for the slope, the transverse deflection, radial stress, and tangential stress, which are developed on the top of the membrane when subjected to a uniform in-plane stress provided that the membrane is circular. Define the terms used in the expressions.
13. How does an MEMS gyroscope work? Define coriolis force and coriolis acceleration. Discuss the compensation techniques used in the process of fabricating the gyroscope. Explain the meaning of DSA.
14. Write short notes on the following
    (a) Shearmode piezoactuator
    (b) Gripping piezoactuator
15. With suitable diagrams, describe in detail the principle of Inchworm Technology.

Chapter 6

Thermal Sensors and Actuators

Objectives

The objective of this chapter is to study the following.

- Thermal basics for MEMS design
- Thermal principle: An introduction
- Thermoelectric effect and Peltier effect
- Thermoresistivity
- Basics of conduction, convection and radiation
- Thermistors, governing equations and their characteristics
- Thermodevices
- Principle of thermocouple and thermopiles
- Peltier heat pump and heat sinks
- Hotwire and microhotplate based thermal flow sensor
- Application of microthermovessels for PCR
- Shape Memory Alloy (SMA) and principles
- Pyroelectricity and application
- U-shaped horizontal and vertical uni- and bidirectional thermal actuators
- Bistable MEMS relays for power system applications
- Chevron (microspring) thermal actuator
- Micromachined thermocouple probe for data storing applications

6.1 INTRODUCTION

Thermal MEMS refers to the microsystems whose functionalities rely on heat transfer phenomenon. Thermal MEMS are categorically divided into three types: thermal sensors, thermal actuators and data storage devices.

When a physical variable to be measured generates an observable thermal profile by the use of solid-state device, such device is said to be a thermal sensor. The physical parameters could be heat radiation, non-radiant heat flux or any other parameter, not necessarily heat flow or temperature. For example, gas

pressures, mass or volume, and fluidic thermal conductivities can be measured using the principle of thermal sensing. Like all other sensors, thermal sensors convert physical variables through a fundamental process called transduction. The transduction phenomenon is mainly observed in three ways, namely thermoelectric effect, thermoresistive effect and the pyroelectric effect. The transduced signal is usually electrical in nature.

Thermoelectric effect implies that a voltage is developed in a loop containing two dissimilar metals when the end junction of the loop is heated. A junction of two dissimilar materials forms a basic thermal device called thermocouple. The opposite of the thermoelectric effect is called Peltier effect, in which current flow causes a temperature difference between the junctions of different metals. The thermoresistive effect refers to a change in the electrical resistance of a semiconductor with temperature. Pyroelectricity is a migration of positive and negative charge to opposite ends of a crystal's polar axis as a result of a change in temperature. The migration of charge is referred to as electric polarization. A more detail thermal principles are briefly described in the Table 6.1

Table 6.1 *Thermal principles*

Name of effect	*Macroscopic description*
Curie temperature	Change to paramagnetism of ferromagnetic material at specified temperature
Electrothermal	Generation of heat in a conductor by electric current
Incandescence	Emission of radiant energy when material is heated
Nemst	Generation of electromagnetic field due to temperature gradient
Peltier	Generation of temperature difference between two junctions when current passes through them
Pyroelectric	Change of polarization due to temperature change
Thermochemical	Change of structure due to temperature
Thermoconductivity	Change of conductivity due to temperature
Thermodielectric	Change of permittivity of a ferroelectric due to temperature
Thermoelectric, Seebeck	Generation of electric potential by a joint of two dissimilar conductors
Thermoluminescence	Emission of radiant energy of certain crystals due to temperature

Source: Uppsala Univ.—http://www.signal.uu.se

Thermal actuators are considered as one of the most important MEMS device, which are able to provide a relatively large force and displacement and are very useful for several applications including electro-optical-communication, micro-assembly, data storage and microtools. Thermally activated important actuation mechanisms are pumping and valving, which are usually employed in the bio-analytical microsystems such as *lab-on-a-chip* (loac) devices. The thermal actuator is useful for switching and positioning applications, as well as for the measurement of mechanical properties of materials. Regardless of their applications, this chapter presents the fundamentals of actuations and the fabrication methods of some important actuators.

Thermal actuation is based on electrothermal energy density transformation which is given by,

$$E = V^2/\rho L^2 \tag{6.1}$$

where, V is the applied voltage, ρ is resistivity and L is the effective length. Note that for some configuration thermal actuation provides higher energy density than piezoelectric and electrostatic

actuation. In order to achieve high sensitivity and responsiveness in the device, the thermal mass of the transducer element has to be kept as low as possible. This is only achieved by using thin film structures made by micromachining techniques.

Alloys such as NiTi, CuZnAl and CuAlNi can change their solid-state phase. The property is called shape memory effect, and can be exploited for actuation applications. These alloys are called Shape Memory Alloys (SMA). After alloying and passing through some basic processing, SMAs can be formed into a particular shape, a coil spring for example, and then set to that shape by a heat treatment. When SMA shapes are cooled, they may be bent, stretched or deformed and then with subsequent heating, but below the heat setting temperature (called memory transformation temperature; MTT), they can recover the deformation. The main advantages of SMAs are that they are bio-compatible. Further, since they have good mechanical properties such as strong and corrosion resistant, SMA can also be applied to diverse fields.

6.2 THERMAL ENERGY BASICS AND HEAT TRANSFER PROCESSES

Energy is always present but not visible! Even though the energy is not visible it can be detected, displayed and used. There are various forms of energies like sound, light, electricity, thermal, etc. Other forms of energy includes gravitational, elastic, chemical, nuclear, and so on. Eventually, there are only two types of energy; kinetic energy and potential energy.

Thermal energy plays an important role in designing MEMS. In particular, it actively takes part in controlling the MEMS actuator. Thermal energy is quantified by temperature. There are several ways of expressing temperature and heat. Temperature is the physical property of a material that signifies whether it is hot or cold. Two elements that are at different temperatures transfer thermal energy from hotter region to colder region. Heat is the transfer of thermal energy within the element. In the molecular scale, the more thermal energy an object has, the faster its molecules move. Thermal energy is connected with kinetic energy of the constituent molecules. An object at a higher temperature means more thermal agitation of its molecules. It is obvious for volume containing higher molecular kinetic energy to pass this energy to volume with lower kinetic energy. The faster moving molecules collide with each other more frequently and as a result of which they need more space, apparently decreasing the density of the molecules. This feature of material at microscale plays important role while designing, studying, simulating and optimizing not only MEMS but also NEMS (Nanoelectromechanical Systems) devices (see Chapter 14) at the primitive stage (called prototyping stage) of micromanufacturing. In summary, it can be said that the temperature corresponds to the quantity, i.e. the amount of thermal energy available in the element and the heat change or heat flow symbolizes the movement of thermal energy from one place to other. Thermal energy can be transferred in three ways, namely conduction, convection and radiation.

Conduction is a process that transfers energy from one molecule to another. The transfer occurs as the adjacent molecules hit against each other. Eventually, a transfer of kinetic energy takes place. Conduction takes place in solids, liquids, and gases, but the process works best in materials that have molecules located close to each other and are said to be compact. This is the reason why metal is a better conductor than plastic or wood. The conduction phenomenon is also called gradient heat transportation. Gradient heat transport depends on three quantities such as the spatial gradient of temperature, the conductivity and the cross-sectional area of the material. The larger the gradient,

conductivity, and/or cross-section, the faster the heat flows. Different materials transfer heat by conduction at different rates. The rate is measured in terms of what is known as thermal conductivity. Mathematically it can be expressed as,

$$t_c = \left(\frac{l}{a}\right) \times \left(\frac{Q}{t_g}\right) \tag{6.2}$$

where, Q is heat flow over time, t_c is thermal conductivity in W/m°C, a and l is the area in m^2 and length in meter of the material, respectively, t_g is the temperature gradient in °C. Thermal resistance is defined as the length over thermal conductance. Thermal conductivity of solid materials such as metals, ceramics, polymers, composites, glass or rubber are measured by thermal conductivity meter.

One can note that metals, which are good thermal conductors, are also good electrical conductors. The thermal and electrical conductivities of metals are directly proportional at a given temperature. However, the experimental study shows that when temperature increases the thermal conductivity increases while the electrical conductivity decreases. The reason behind this observable fact is that both the heat and electrical transport processes engage free electrons in the metal. In thermal process, the conductivity increases since for transport of thermal energy average particle velocity increases with temperature. However, in the latter case the conductivity decreases as the engagement (vibration, collisions, etc.) causes diversion of the free electrons from forward transport of charge. A useful conclusion can be derived in terms of defining the ratio of these two conductivities. The ratio is known as Wiedemann-Franz Law.

$$\frac{t_c}{\sigma} = Lt \tag{6.3}$$

where, L is called Lorenz number, t is the temperature, σ is the electrical conductivity.

When a material is heated through a temperature range, it expands. As explained earlier, the flow of heat makes it possible to increase the kinetic energy of atoms, thus causing more distance between the atoms. With regard to such expansion materials are classified on the basis of whether they are isotropic or anisotropic. In an isotropic material, the expansion occurs uniformly in all dimensions, whereas non-uniform expansion occurs in case of anisotropic materials.

If an object of length L_1 is heated through a temperature change ΔT, then the original length of the object will change to $L_1 + \Delta L$. The change in length ΔL is proportional to the original length L_1 and to the change in temperature. Mathematically, this can be expressed as,

$$\Delta L = \alpha L_1 \Delta T \tag{6.4}$$

where, α is known as the coefficient of linear expansion. Many materials are not isotropic and therefore have a different values of α along the axis upon which the expansion is measured. Immediately we can also define the coefficients of *area* and *volume* expansions, as follows.

$$\Delta A = \beta A_1 \Delta T \tag{6.5}$$

$$\Delta V = \gamma V_1 \Delta T \tag{6.6}$$

where, β and γ are known as the coefficient of *area* and *volume* expansion, respectively.

Besides thermal conductivity and thermal coefficient of expansion, there is another thermal constant that illustrates the property of material called heat capacity of the material. The heat capacity of a material is the amount of heat required to change its temperature by 1 °C. A large measure of substance will have a proportionally large heat capacity. At constant pressure the heat capacity can be expressed as,

$$C_p = \left(\frac{dQ}{dT}\right)_p = T\left(\frac{ds}{dT}\right)_p \tag{6.7}$$

where $(dQ/dT)_p$ is the change in heat with temperature, T is the temperature, and S is called entropy, a measure of a system. A more useful quantity is the specific heat, also called specific heat capacity, which is the amount of heat required to change the temperature of one unit of mass of a substance by one degree.

Convection is another process in which the movement of liquid such as water or gas causes heat transfer. This movement of a mass of heated water or gas is sometimes called current. Radiation is the transfer of heat by electromagnetic waves.

There exists another concept, which appears as the principle of thermal energy storage (TES) that deals with the storing of energy either by cooling, solidifying, heating, melting, or vaporizing a material. The transition from solid to liquid and vice versa or from liquid to vapor and vice versa with no change in temperature is an example of heat storage. The energy becomes available as heat when the transitive process is reversed. Storage capability and efficiency depends on the *specific heat* as well as the density of the storage material. The specific heat of solidification, fusion, or vaporization and the temperature at which the phase change occurs must be known before using the material for MEMS applications.

6.3 THERMISTORS

Temperature is among the most important parameters to be measured for many automation, control and consumer product applications. The common types of temperature transducers in use today are thermal resistors or *thermistors*. A thermistor is a piece of sintered semiconductor material, which exhibits a relatively large change in resistance proportional to a change in temperature. Thermistor possesses negative temperature coefficient of resistance, i.e. the resistance of the piece decreases with increase in temperature. The relationship between resistance and temperature for thermistors is expressed in Eq. 6.8.

$$r_t = r_0 e^{\rho\left(\frac{1}{t} - \frac{1}{t_0}\right)} \tag{6.8}$$

where r_t is resistance in Ohm at temperature t, r_0 is the resistance of the thermistor at a reference temperature t_0, ρ is the property of the semiconductor material.

For precise temperature measurement, thermistors are considered to be versatile, rugged, and cost effective. High sensitivity, stability, and accuracy, make thermistors the most popular compared to other temperature sensing devices. They typically work over a relatively small temperature range and can produce accurate equivalence. The stability of the thermistor is expressed using a term called *drift*. The drift is quantified as a change in resistance that occurs at a given exposure temperature for a certain length of time. The drift increases as the limit of the temperature and the duration of exposure increases. Adverse exposure degrades the performance of the thermistor.

Thermistors can also be designed from metals. Metal thermistors have positive temperature coefficients of resistance (Refer to the table in Appendix B). These are rather called Resistance Temperature Detectors (RTD). RTDs are wire wound and thin film types that work on the physical principle what is known as the temperature coefficient of electrical resistance of metals. The equivalence is nearly linear over a wide range of temperatures and has response times in the order of a fraction of a second. Like Light Dependent Register (LDR) they require an electrical current to produce

a voltage drop that can then be calibrated and displayed. The relationship between resistance and temperature for metal type (RTDs) thermistors are expressed in Eq. 6.9.

$$r_t = r_0(1 + \alpha(t - t_0)) \tag{6.9}$$

where, r_t is the resistance in Ohm at temperature t, r_0 is the resistance of the RTD at a reference temperature t_0, α (°C^{-1}) is temperature coefficient of the material. These transducers are made from various compositions of the metal oxides of manganese, nickel, cobalt, copper and iron.

Figure 6.1 shows temperature versus resistance curve of typical semiconductor-based thermistors and RTDs, satisfying Eq. 6.8 and Eq. 6.9, respectively.

Among many configurations, *beads*, *discs*, and *chips* are the most widely used thermistor types, due to the reason that these configurations are typically more rugged and better able to handle mechanical shock and vibration. Thermistors are manufactured by the *sintering* process, which consists of (i) the preparation of the metal oxide, (ii) milling and blending, (iii) heat-treatment, (iv) addition of electrical contacts, (v) assembly into a device, and (vi) protective coating. Specification of a typical thermistor is 10000 Ohms at 25 °C, with a resistance tolerance of ±10%.

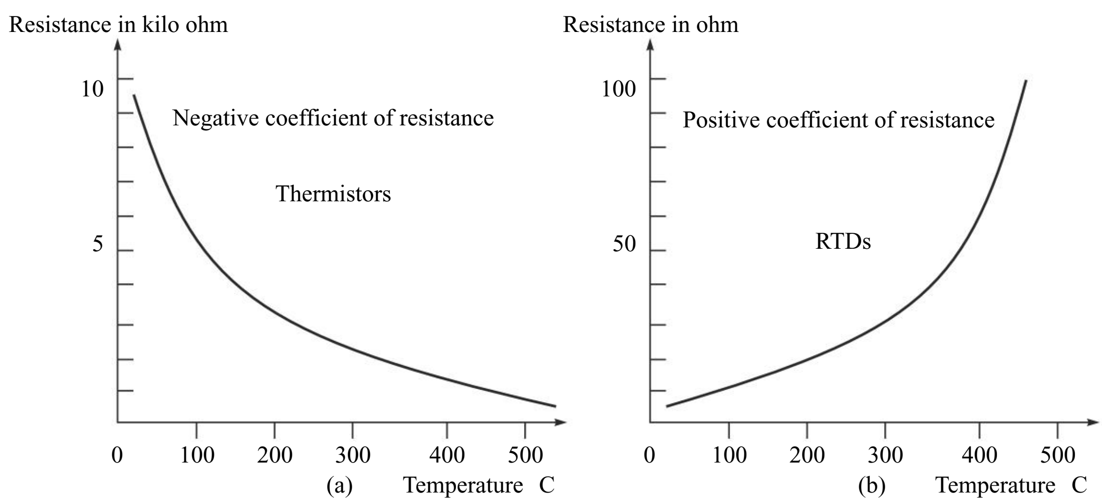


Fig. 6.1 *(a) Temperature-resistance curve of typical thermistors, (b) of RTDs*

6.4 THERMODEVICES

Semiconductor diodes and transistors specially designed to meet the requirements of the temperature measurement are individually called *thermodiodes* and *thermotransistors* respectively and commonly called *thermodevices*. Themodevices are more sensitive to the temperature variations as compared to other devices. The characteristic and design features are such that their junctions are more sensitive to temperature. The principle of operation of thermodevices is that if temperature increases, the thermal energy makes it possible to knock out the hole-electrons pairs from the valence band of the semiconductor materials. These knocked out charge carriers, acquiring higher energy then enter into the conduction band, which in effect affects the junction voltage. Identical situation occurs at the emitter-base junction of the thermotransistors. Thermotransistors are more sensitive as compared to thermodiodes, since the output voltage in this case is the amplified version of the voltage developed at

the emitter-base junction. The voltage across the diode junction is related with the temperature as follows,

$$V = \ln\left(\frac{I}{I_0} + 1\right)\frac{kT}{q} \tag{6.10}$$

where I is the current through the diode, I_0 is called reverse saturation current, k is Boltzmann constant (1.38×10^{-23} Joules/Kelvin), T is the temperature in degree Kelvin and q is the charge of the electron. Temperature sensors are available as Integrated Circuit (IC), in which thermodevice as well as signal conditioning circuits such as amplifying circuit and analog-to-digital converter (ADC) are built in to the device.

6.5 THERMOCOUPLE

A thermocouple is a transducer that is also employed for the measurement of temperature. It consists of two dissimilar metals, joined together at one end, called thermocouple junction. When the temperature of this junction is different from the temperature of other parts of the metals an EMF (Electromotive Force) is generated as shown in Fig. 6.2. The phenomenon is called *Seebeck effect* in honor of Thomas Seebeck, who first noticed the phenomena in 1821. The EMF produced is known as *Seebeck* EMF or simply *thermovoltage*, V_{th}. It is important to note that different metal combinations have a different thermovoltage level. If points '*a*' and '*b*' are connected, a current will flow through the circuit. The loop thus formed in this way is called *thermoelectric circuit*. The variation of thermovoltage in response to the temperature variation is measured, calibrated and interpreted as temperature.

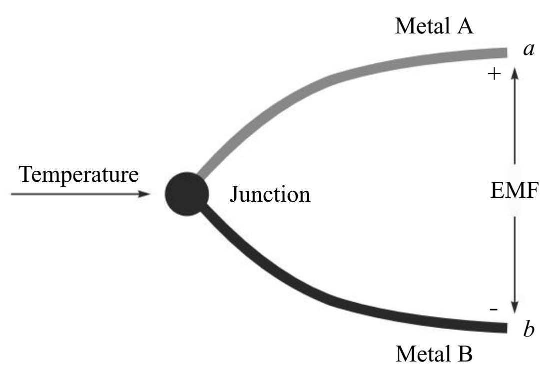


Fig. 6.2 *A schematic diagram of a typical thermocouple*

Using only one kind of metal no temperature can be detected. It is evident that when a strip of metal is heated at one end relative to the other the electrons at this end (hot) will acquire thermal energy. The electrons from the hot end will diffuse to the other end and get accumulated. In effect, a tiny amount of electric field is established and a situation will arrive when there will be no net further diffusion. The developed electrostatic field will make it possible to repel the electrons. Although, there will be no net flow of electrons across a cross section, the velocity of electrons on two sides of the cross section remains different. The velocity at hotter side is greater as compared to the colder side. Because of velocity differences a continuous transfer of heat is ensured. This phenomenon is referred to as *thermal conduction*, which simply means that a temperature gradient has to be established albeit there involves no net flow of electron even after equilibrium is set up. Under these circumstances, to measure the amount of electric field the measuring device must be connected to either side of the strip. If the measuring device is made up of the same material as that of strip, then the strip and the measuring device can be considered as not two elements rather a single type thus establishing a similar temperature gradient in the arrangement. Therefore, under these conditions, no net EMF would be detected. To

measure the EMF, it is desirable that the thermoelectric circuit should have at least two different materials.

The relationship between the temperature and the produced thermovoltage is expressed in Eq. 6.11.

$$V_{th} = \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \dots + \alpha_n t^n \tag{6.11}$$

where V_{th} is thermovoltage, t is the temperature to be measured and α are the thermocouple coefficients, which depend upon the materials used and are temperature independent. Note that the relationship is not linear. Considering more number of terms in the above power series can minimize the *nonlinearity error*. In most cases, the best result is obtained with tolerable accuracy if approximately three numbers of terms are taken. The term α_1 is known as *Seebeck coefficient*. Materials having good level of Seebeck coefficient are always preferred. Remember that the coefficients are temperature independent and material dependent. For best linearity the selection of materials plays important role. The important factor in determining the materials is that the produced EMF should be linear with respect to the temperature.

Practically, a thermocouple loop looks like a circuit as shown in Fig. 6.3. There are two junctions J-1 and J-2. Keep in mind that it is impossible to connect the thermocouple to any temperature-measuring device without constructing another junction. The junction J-1 is the junction responsible for the measurement of temperature with respect to the junction J-2, which is called cold junction or reference junction. J-1 is called hot junction. The arrangement is fairly a differential device and the output is related to the difference in temperature between the hot and the cold junction. The cold junction is held at a constant temperature. This essentially allows for accurate calibration of the temperature of the hot junction.

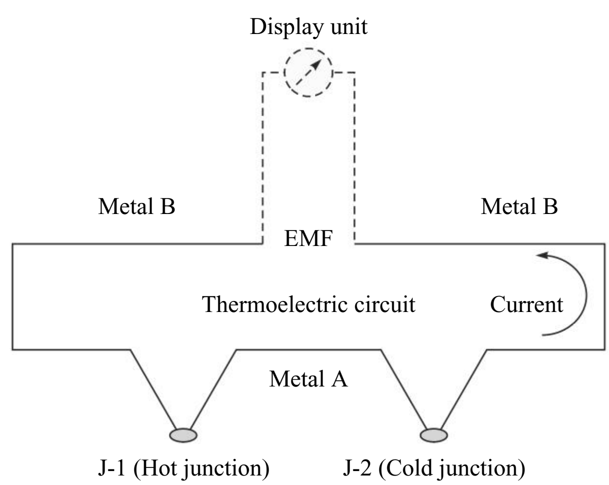


Fig. 6.3 *Thermoelectric circuit*

Fundamentally, any attempt to join any two dissimilar metals express a thermocouple. As expected, if a thermocouple-based temperature sensor is connected with the signal conditioning circuitry, more or less the entire integrated system forms additional thermocouples along with the main thermocouple, thus producing additional thermovoltages, which are superimposed onto the base thermovoltage produced by the main thermocouple. However, it is a fact that the other metals in the circuit have no

effect on the base thermovoltage as long as the temperature of the new junctions (because of connections) are kept at the same level.

The important parameters to be considered in selecting thermocouples for desired applications are sensitivity or *Seebeck* coefficient, temperature range, and resistance to corrosion, vibration and strain.

Alloys are usually used for the construction of thermocouple. Ageing causes variations in the alloy composition for which the performance deteriorates with time. Another source of error is moisture. This is the reason why the thermocouples are supplied with the protecting case. Several thermocouple types have been emerged as standard transducers because of remarkable features and qualities in terms of linearity and sensitivity. Sensitivity is defined as a ratio of voltage to temperature. Table 6.2 provides information about some of the standardized thermocouple types and their characteristics.

Table 6.2 *Standardized designated thermocouples*

Sl. No.	*Designation*	*Composition (terminal)*	*Range*	*Sensitivity ($\mu V/°C$)*
1	B	Platinum 30% Rhodium (+), Platinum 6% Rhodium (−)	1300–1700	3
2	C	W5Re Tungsten 5% Rhenium (+), W26Re Tungsten 26% Rhenium (−)	1650–2315	-
3	E	Chromel (+), Constantan (−)	−100–900	64
4	J	Iron (+), Constantan (−)	−150–800	52
5	K	Chromel (+), Alumel (−)	−200–1200	40
6	N	Nicrosil (+), Nisil (−)	500–1300	29
7	R	Platinum 13% Rhodium (+), Platinum (−)	0–1500	6
8	S	Platinum 10% Rhodium (+), Platinum (−)	500–1400	6
9	T	Copper (+), Constantan (−)	−200–350	44

Sometimes, the output voltage of a single thermocouple is too small to detect. The signal strength can be improved by connecting several thermocouples in series. The series connection of thermocouples constitute a thermopile.

6.6 MICROMACHINED THERMOCOUPLE PROBE

Thermal probes are devices specifically designed for precision monitoring of temperature in a variety of applications, including photothermal absorption, spectroscopy, subsurface imaging, photolithography research, micro-calorimetric analysis and scanning microscopy.

Thermal probes are manufactured based on various design techniques such as Schottky diodes, thin-film bolometer, coated wire thermocouple and micromachined thin film thermocouple (TFTC). Out of these, TFTC probes are MEMS devices. TFTC devices are attractive because of the following factors.

- Fabrication is relatively simpler
- More localized temperature sensing
- Bias current is extremely small. In non-MEMS based devices, high value of bias current could cause unnecessary heating degrading the efficiency.

The microprobe is an essential element in the applications of scanning probe microscopy (SPM). SPM is a technique that is used to study the properties of surfaces at the atomic level. The surface

properties are alignment, roughness and elevations. SPM is useful for measuring surfaces on a fine scale down to the level of molecules. The technologies share the concept of scanning an extremely sharp tip across the object surface. The radius of curvature of the tip ranges from 5−50 nm. The tip is mounted on a flexible cantilever based probe, allowing the tip to follow the surface profile. The interaction between the sharp tip and surface provides 3-D topographic image of surface. The topography is defined in terms of relative elevations of the surface profile. Scanning probe equipped with scanning tip, overhanging at the edge of the substrate makes it possible to map the alignment and microelevations of the sample surface.

The scanning tip integrated microprobe is simply a cantilever upon which a thin-film thermocouple resides. Figure 6.4 illustrates the schematic layout of a typical polyimide-based thermal microprobe. Two layers of polyimide is partially released from the surface of the wafer and turned over so as to enable it to expose the scanning tip outward. The polyimide material based probe is flexible and robust.

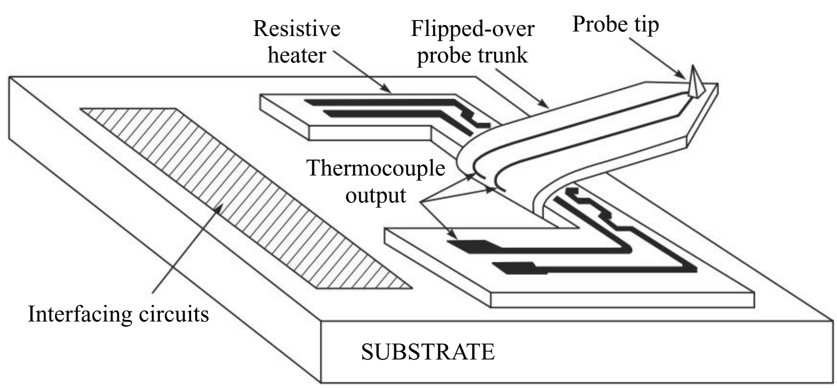


Fig. 6.4 *Schematic of the thermal probe*
Source: Li et al., *J. of MEMS,* 10/1, 2001

The principle of operation of the thermal probe is that a temperature bias is provided between the thermocouple junction and the sample. The probe itself is a thin film resistor that delivers a temperature bias with respect to the sample upon which the measurement will be carried out. At the vicinity of the tip the hot junction of the thermocouple is present (TC_2) whereas other one (TC_1) is at the unreleased base region (see Fig. 6.4 and Fig. 6.5). The difference in temperature causes thermal conduction across the air gap and along the probe trunk. A significant fraction of the temperature bias drops across the air gap and the remainder drops along the probe trunk, which is the span between two thermocouple junctions. Since there is a temperature difference the probe can be used to read out the signal. The read out signal will vary since it responds to the variations in the air gap between the tip and the sample surface. By employing two junctions, the common mode errors (CME) which results from global temperature drifts or other effects can be cancelled out.

Let us define R_{sh} and C_{sh} as the distributed thermal resistance and capacitance (Fig. 6.5) of the probe trunk, R_g as the resistance of the air gap. For a given R_g, the temperature drop along the trunk length is highest when R_g becomes equal to R_{sh} (recall the maximum power transfer theorem). Note that the response speed of the probe can be increased by reducing the value of the product of distributed thermal resistance and capacitance of the probe trunk.

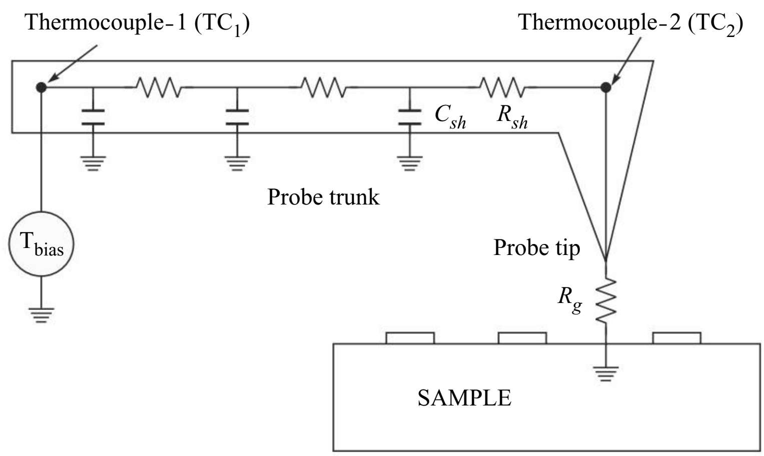


Fig. 6.5 *Thermal conduction model of the probe including sample*
Source: Li et al., J. of MEMS, 10/1, 2001

The fabrication process in making the thermal microprobe is as follows (Figs. 6.6(a)–(f)). (100) oriented Si substrate wafers can be used as substrate.

- An oxide masked anisotropic wet KOH etch is first used to define a pyramidal notch for the tip (Fig. 6.6(a)).
- A 2 μm thick sacrificial layer of Ti is deposited and patterned using a photoresist masked wet etch (Fig. 6.6(b)).
- The first polyimide layer is then spun on, cured, and patterned, removing it from the field regions. An aluminium hard mask is used with a dry reactive ion etch of O_2 and SF_6 to pattern the polyimide (Fig. 6.6(c)).

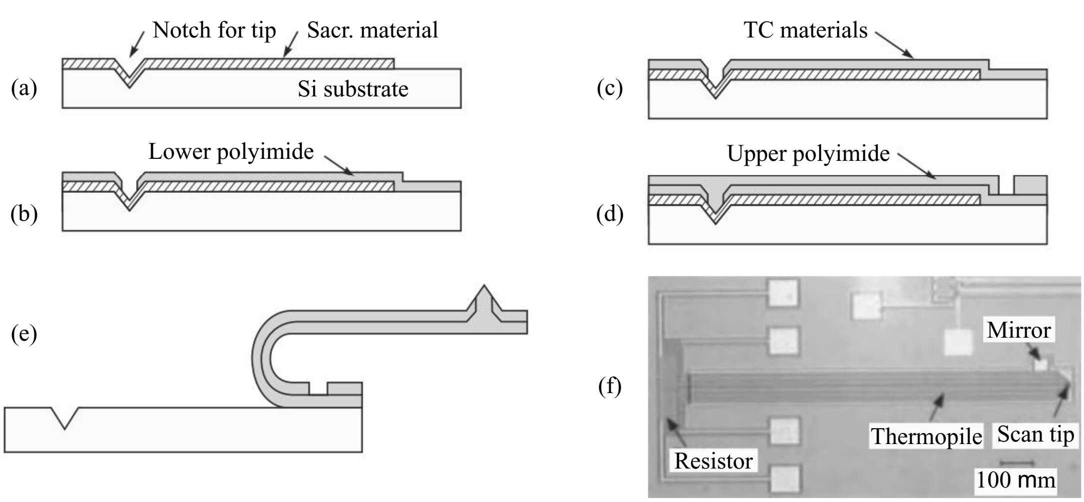


Fig. 6.6 *(a)–(e) Fabrication process, (f) Micrographs of fabricated thermal probes immediately before release*
Source: Li et al, *J. of MEMS,* 10/1, 2001.

- The thermocouple (TC) metals are deposited and patterned (Fig. 6.6(d)).
- Finally, the probe is released (Fig. 6.6(e)).

The typical dimensions are 200–1000 μm long, 40–120 μm wide, and of varying thickness. The probe can be used for force and tunneling microscopy.

6.7 PELTIER EFFECT HEAT PUMPS

In 1834, the French physicist Jean Peltier discovered an effect inverse to the Seebeck effect, which is known as Peltier effect. The effect implies that a thermocouple can function in reverse.

The Peltier effect states that if an electric current I is passed across the junction of two different conductors (e.g. through a thermocouple—Fig. 6.7(a)), a change in temperature at the junction is observed. The quantity of heat liberated per unit time is proportional to the current. Since the density of electrons in the two materials is usually not the same the phenomenon produces heat at the rate given by,

$$\pm Q = (\Theta_a - \Theta_b) \cdot I \tag{6.12}$$

where, Q is the rate at which the heat is produced; Θ_a and Θ_b are called the Peltier coefficients of the two dissimilar materials A and B. The sign of Q can either be positive or negative. A negative sign implies cooling of the junction. The extent of the temperature change depends on the conducting metals, the nature of change (i.e. rise or fall in temperature) and the direction of current flow.

The relationship between the Peltier coefficient and Seebeck coefficient is given by,

$$\Theta = \alpha_s \cdot T \tag{6.13}$$

where, T is the absolute temperature.

When electric current I flows in a homogeneous conductor in the direction of a temperature gradient defined by dT/dx, heat will be released or absorbed depending on the material. This is known as Thomson effect and mathematically it can be expressed as,

$$Q = \alpha_t \cdot I \cdot \frac{dT}{dx} \tag{6.14}$$

where, α_t is called Thomson coefficient. The direction in which the heat flows depends on the sign of the Thomson coefficient, the direction in which the current flows and the direction of the temperature gradient.

The Peltier effect is very useful in manufacturing devices for heat pumping applications. Such a pumping device is called heat pump or Peltier Module (PM) that can be manufactured using MEMS technology. The module is designed by placing p and n type semiconducting materials alternatively in series as shown in Fig. 6.7(b). High electrical low thermal conductive semiconductor materials are desirable for heat pump. Bismuth telluride is a preferred material. The p and n type materials are connected electrically in series but thermally in parallel. At each end they are soldered to copper connecting strips. More than 100 p–n junctions can be placed in series. The module is electrically insulated from the surface by ceramic faceplate.

When current is passed through the module one side becomes cold and other side becomes hot. The effect is reversible by passing the current in other direction. Such a configuration called a *heat pump* because the heat is given out or pumped out of the cold surface and is deposited on other surface (Fig. 6.7(c)). The combination of many pairs of p- and n-semiconductors allows creating powerful heat pumping unit (also cooling unit). The power of the module depends on its pn bar density. Within the module another component called heatsink is attached. The function of the heatsink is to dissipate the

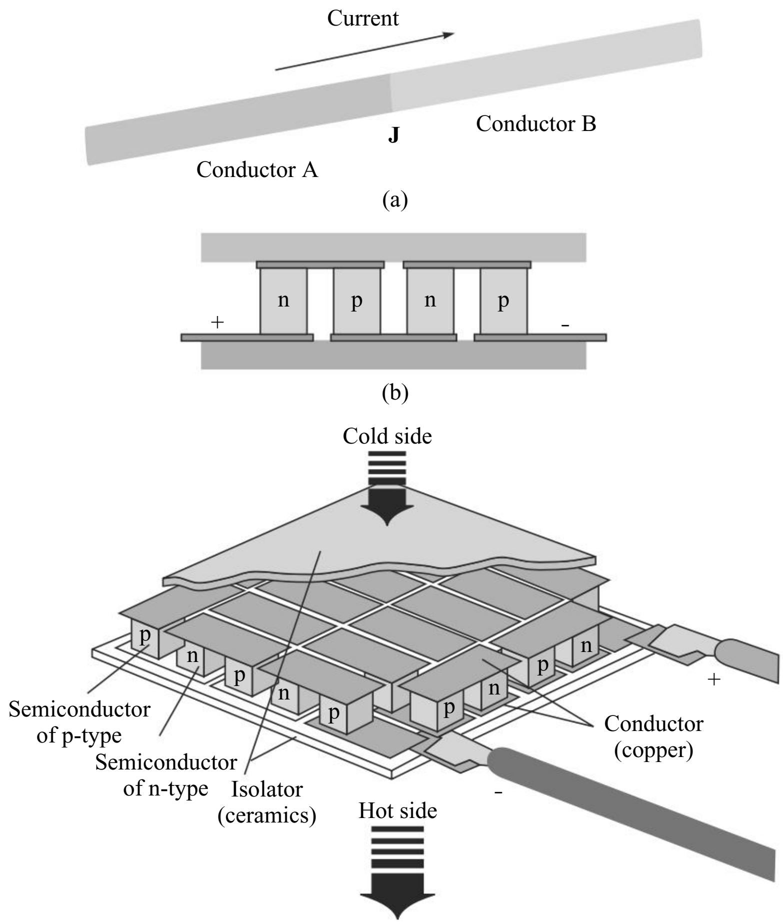


Fig. 6.7 *(a) The Peltier effect; (b) Usage of semiconductors of p- and n-type in thermoelectric coolers; (c) Structure of a Peltier module*
Source: E. Rudometov and V Rudometov, bhv Publishing

pumped out heat so that the cooling system can work properly without rise in temperature above a tolerable level. In practice, we must choose a heatsink, which will absorb the total waste heat from the module. An ideal heatsink would be capable of absorbing an infinite amount of heat without rising in temperature. Heatsinks are rated in ^{o}C/Watt.

Heat dissipation is a big problem in many kinds of computing devices. Since the chip circuitry continues to reduce in size and run ever faster, CPUs and other chips are giving out more heats. Peltier modules (PM) are useful in these fields. A PM containing more than 100 (Bi, Sb)$_2$Te$_3$ n-type and p-type thermoelectric elements bridged by metal interconnects with overall dimension of 20 μm in height and 60 μm in diameter is already available in the technology marketplace. Such a device also finds technological importance in other applications in the sense that it can be used for precise thermal control when operating as a cooler, and for portable power when operating as a micro power generator (Fig. 6.8).

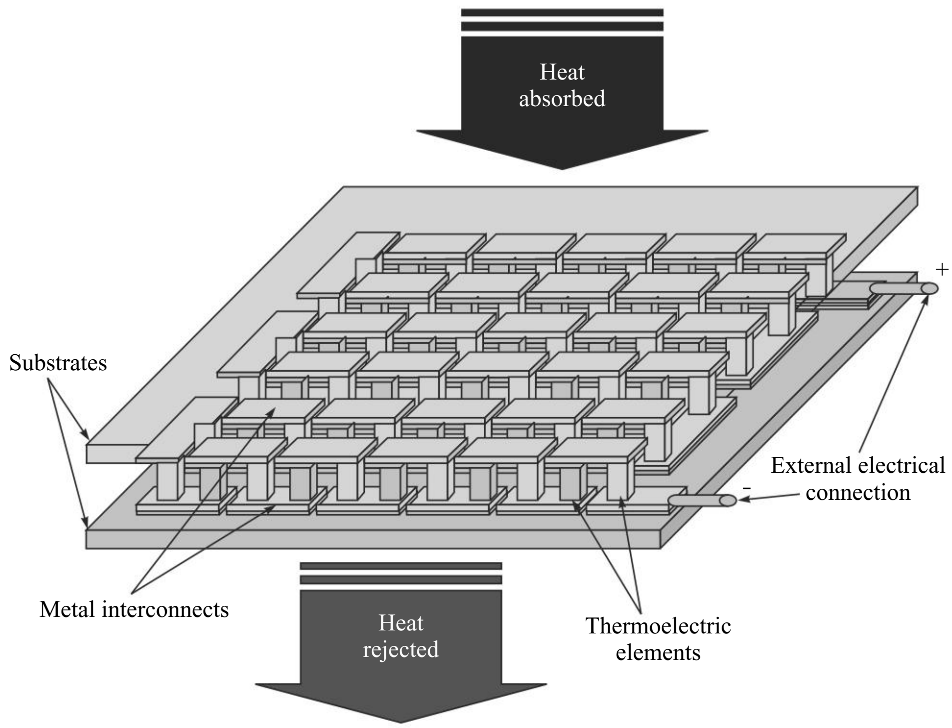


Fig. 6.8 *A Peltier module conformant thermal controller cum micro power generator*
Source: Nature Materials, 2, pp. 528–531, 2003.

6.8 THERMAL FLOW SENSORS

MEMS thermal flow sensors are designed to precisely measure air flow rate of cooling systems, burning appliances and ventilators. The MEMS flow sensor measures flow rates from 0 to 100 m/s and can deal with flow rate changes up to 80 Hz. The MEMS flow sensor can also be used in a pulsating flow and in a vibrating environment. They can withstand the impact of small particles. Many of the flow sensors (Fig. 6.9) adopt thermo-dynamic dispersion method (TDDM) as precision measurement is becoming momentous. The principle behind the method is the Peltier effect.

6.8.1 Types of Flow Sensors Based on Working Principle

The efficient operation of systems and equipment in modern technical processes including automation, control, instrumentation, materials processing and discrete manufacturing plants frequently depends on the accurate flow measurement of gases and liquids. The gas to be measured is allowed to flow across the surface of the sensor. There is a heater that heats the sensor surface. The passage of the gas cools the surface and the resulting surface temperature is measured. There are mainly two ways by which the flow rate of gases and liquids can be measured. One way of utilizing the Peltier phenomenon is by measuring the temperature of a fluid as it enters the sensor, and then as it leaves the sensor having been passed over a heating resistor. When there is no flow, temperature distribution concentrated around the heater is uniform, but when there is a flow, temperature on the side of the heater facing the flow cools

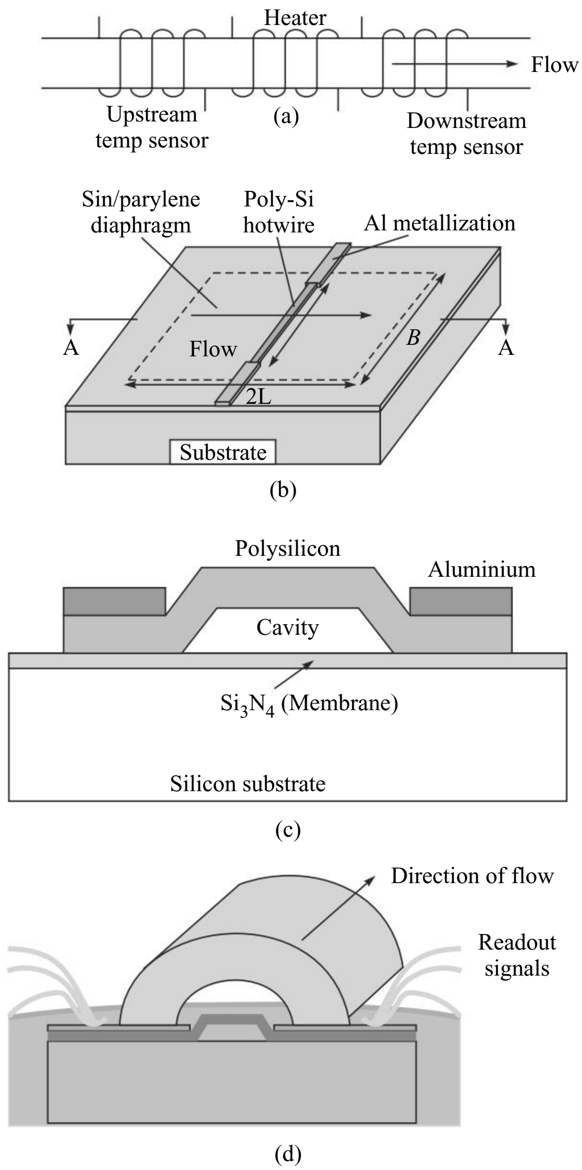


Fig. 6.9 *(a) A flow sensor showing heater resistor, upstream and downstream temperature sensor (b) The schematic diagram showing the heater (called hotwire) and the direction of flow (c) The suspended hotwire over the cavity (d) A view of the covered hotwire and the direction of flow of gases*

(upstream), the side away from the flow (downstream) warms up (Fig. 6.9(a)). The temperature difference between the upstream and downstream temperature sensors becomes proportional to the mass flow rate. Such types of sensor are called *differential sensor*.

In another method it is required to maintain the sensor surface at a constant temperature using heating resistors and with feedback loop that measures the amount of power required to maintain the

temperature in order to achieve thermal equilibrium that was collapsed because of gas flow. In this scenario the required power is proportional to the mass flow rate of material over the sensor. These sensors are called *equilibrium sensors*. An equilibrium sensor does not have upstream and downstream temperature measuring devices since they operate in constant temperature mode. Feedback control is used to maintain the hotwire at a fixed temperature and only the electric power consumed in the hotwire is to be measured. In this case the heat transfer from the hotwire to the fluid flow is dependent upon the thermal characteristics.

6.8.2 Schematic of Differential Type Flow Sensor

A schematic of a differential type flow sensor is illustrated in Fig. 6.9(b). The sensor has a heater called hotwire which is simply a resistor made up of a polysilicon thin-film lying on a silicon nitride or Parylene membrane. The membrane is suspended from the supporting substrate by a vacuum or air cavity. Sometimes the hotwire is structured in such a way as to ensure that it forms a cavity-like opening instead of simply lying over the membrane (Fig. 6.9(c)). In this case development of another cavity is not required. In order to heat the hotwire an electric current is passed via aluminium metallization through the resistor as shown. To measure the temperature due to fluid flow, temperature sensors are placed on both the sides of the heater (shown in Fig. 6.9(a)) constituting a TDDM based differential sensor. In many designs multiple temperature sensors are integrated at both the ends and the read out is taken statistically. Such type of measurement technique is referred to as sensor fusion technique. Finally, the whole unit is covered (Fig. 6.9(d)) in order to protect it from external intervention. The flow sensor can also detect flow directions. Direction sensor configuration comprises multiple heaters and temperature sensors.

6.8.3 Thermal Isolation

Accurate flow measurement in systems and equipment is not only critical for achieving the necessary temperatures for operation, but also to operate as efficiently as possible and reduce energy consumption. The main concern in all types of thermal microsensors is the thermal isolation on the sensing elements. High degree of thermal isolation is desirable for these relatively high-temperature microdevices. Both types of sensors described above, are designed to incorporate thermal isolation scheme. Thermal isolation improves measurement sensitivity and efficiency. The basis of thermal isolation design is based on fabrication of a thin and narrow polysilicon hotwire over an air cavity. Further isolation is achieved by developing a thick membrane on the substrate. A scanning electron microscope (SEM) picture showing the top view of the sensor based on the porous silicon over air cavity technology is shown in Fig. 6.10. The output of the sensor is a differential measurement of the voltage developed at each thermopile under flow.

6.8.4 Thermal Flow Sensing Array

Flow sensors based on thermal operating principles described above have long been popular for their ease of use and fabrication methods. These sensors have several applications. Sensors for use with biological applications, however, require biocompatibility and low operating temperatures. Parylene and platinum are reasonably biocompatible. For low temperature operation the sensing system entails array configuration. The arrayed device (Fig. 6.11) also operates based on the flow rate dependent convective heat transfer from a heated sensing element to passing fluid. The arrayed sensor is capable of sensing flows as low as 0.5 μL/min. Flow rate is measured by tracking a heat pulse applied to a heater at a sensor downstream. The detected signal measured in terms of resistance change over time.

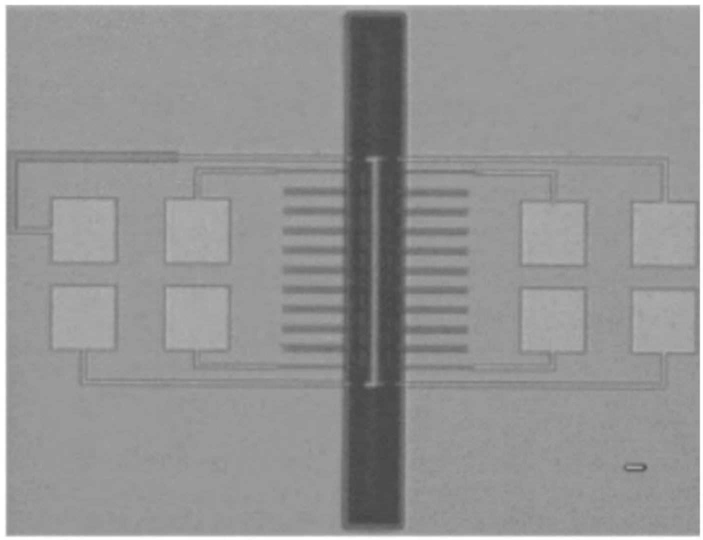


Fig. 6.10 *Top-view picture of a typical thermal flow sensor based on the porous silicon over air cavity technology*

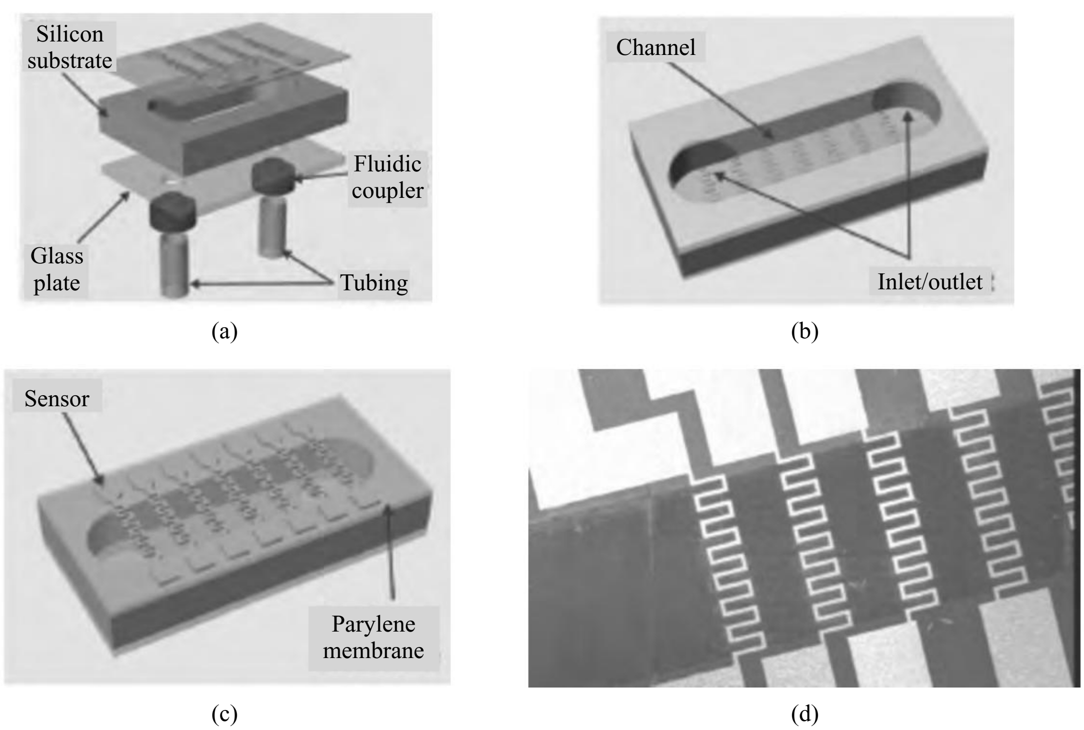
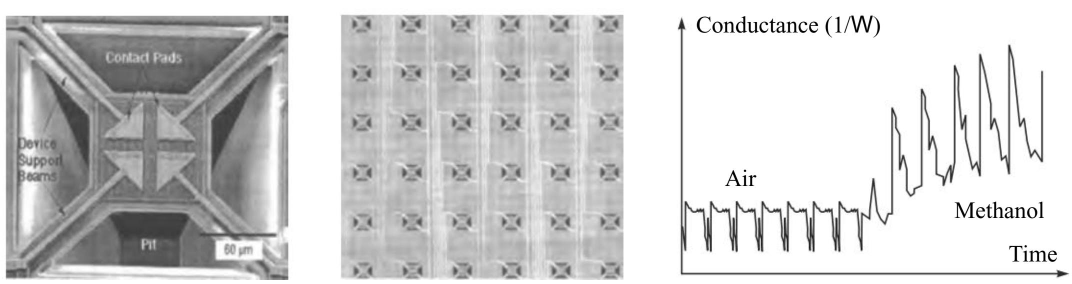


Fig. 6.11 *(a) 3D Views of sensor configuration showing silicon substrate, glass plate fluidic coupler, (b) The channel (Substrate), (c) A closer view of the channel and sensing resistor (heater), (d) More closer view*
Source: Meng and Tai, California Inst. of Tech., *Transducers*, 2003

6.9 MICROHOTPLATE GAS SENSORS

There is another class of thermal flow sensor whose design is based on *microhotplate* instead of hotwire. As name indicates the heating element is plate like membrane. While the principle of operation remains same, the constructional design with regard to thermal isolation is somewhat different. The microhotplate-based flow sensors are highly efficient because of improved isolation. In this case the cavity is formed under another silicon porous membrane. Even lower thermal conductivity of air compared to that of porous silicon assures a superior local thermal isolation on top of the porous membrane. The thermal isolation is achieved because of lower thermal conductivity of porous silicon. The fabrication of various integrated thermal sensors employing porous silicon thermal isolation method is a very successful technology.

A semiconducting film is basically deposited on microhotplate platform that actually detects gas species instead of fluid flow. For instance, SnO_2 can detect hydrogen and methanol. Microhotplate is a micromachined platform with integrated temperature sensing and actuation, mostly used for closed-loop thermal control. Typical dimension of thermal platform has lateral dimensions of less than 100 micrometer, and is suspended over a bulk-etched cavity. A micrograph of a typical microhotplate element is shown in Fig. 6.12(a). The microhotplate gas sensor is actually an array of microhotplate elements (Fig. 6.12(b)).



(a) (b) (c)

Fig. 6.12 *A microhotplate element that can be used for the detection of gas such as air and methanol, (b) An array of microhotplate elements constituting a gas sensor, (c) Sensor output in air and methanol*
Source: DeVoe, *IEEE T on Components and Packaging Tech.* 25/4, 2003.

The physical architecture of each microhotplate consists of a polysilicon heater, a thermoresistive film for temperature measurement, and a semiconducting film, which exhibits a change in conductance with the adsorption of chemical species. With aluminium as the temperature sensing film, the microhotplates can operate up to 500 ^{o}C. Thermal response for 100-micron wide microhotplate is measured around 0.6 ms, with a thermal efficiency of 8 C/mW. Figure 6.12(c) illustrates the response of a single microhotplate in gas like air and methanol. Each peak corresponds to the change in conductivity as the hotplate is rapidly cycled between two extreme temperature points such as room temperature and 500 ^{o}C. Note that the differences in the adsorption kinetics of the gas as a function of the hotplate temperature cause the shape of the conductance response to vary. The signal processing circuitry can be used to identify the gas.

The process developed for porous silicon over cavity formation is a two-step electrochemical process. A porous silicon layer is initially formed on a predefined area on the substrate by anodization of monocrystalline silicon through an appropriate masking layer. Subsequently, by increasing the current

density a cavity is formed underneath the porous layer by electropolishing of silicon. The integration of the gas flow sensor on the suspended porous silicon membrane is a three-mask process.

- TEOS (Tetra-ethyl-ortho-silicate) oxide is first deposited on top of the membrane for electrical isolation while a 500 nm polysilicon layer is then deposited on top of the TEOS layer by low pressure chemical vapor deposition (LPCVD).
- The polysilicon layer is subsequently implanted by boron and patterned in order to form a heater on top of the membrane and the first part of two sets of poly-Si/Al thermopiles at each side of the heater. The 'hot' contact of the thermopiles lies on the membrane while the 'cold' one lies on the silicon substrate.
- The final processing step is aluminium deposition (500 nm thick) and Al patterning in order to form the second part of the thermopiles and the necessary metal pads.

6.10 MEMS THERMOVESSELS

Rigorous thermal analysis and experiments that assist in designing thermo-compatible MEMS structures are in progress. Some effort has been made in the field of liquid-crystal thermometry of micromachined silicon arrays for DNA replication. DNA replication means the use of existing DNA as a pattern or model for the analysis and synthesis of new DNA strands. The replication is achieved through what is known as polymerase chain reaction (PCR) process. PCR requires accurate cycling of the liquid sample temperature with operational range between 55 and 95 degree Celsius. PCR that makes use of micromachined structures, called microthermovessels (MTV), is shown in Fig. 6.13. The miniaturized vessels can utilize less reagent and sample volumes. These are the vital requirements for PCR system that handles liquid samples of the order of micovolumes. The MTV array assures uniformity as far as temperature and cycle timing is concerned. One important requirement is that the vessels should be bio and thermo compatible. Suitable materials are chosen in order to achieve biocompatibility and to establish temperature uniformity within the vessel. It is also required to look into the desired thermal *time constant* of the vessel array. To control the reaction, precise distribution of the temperature in the reacting liquid is essential. For this accurate measurement of temperature distribution is also necessary. The vessel array for DNA PCR is therefore integrated with thermometry circuitry and systems. While the integrated system is useful for characterizing the PCR process through supporting feedback control, it also stabilizes the reaction activities. The design architecture is not only useful for DNA analysis and testing but can also be used for most of the micro fluidic applications (Refer Chapter 10).

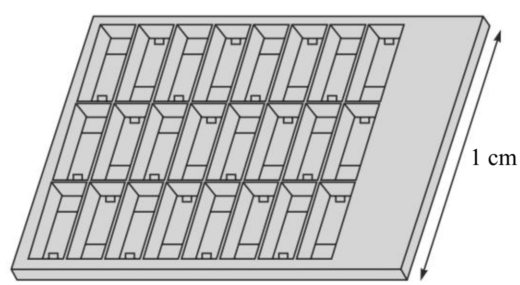


Fig. 6.13 *Micromachined thermovessel array for thermal processing of DNA using PCR*
Source: Applied Biosystems.

6.11 PYROELECTRICITY

Pyroelectricity is a temperature dependent electrical polarization phenomenon. The polarization within the material implies that below a temperature known as the Curie point, crystalline material or

ferroelectric materials (e.g. lithium tantalite) exhibit a large spontaneous electrical polarization in response to a temperature change. Materials, which possess this property, are called pyroelectric material. The change in polarization is observed as an electrical voltage signal if electrodes are placed on opposite faces of a thin slice of the material. The design is a simple form of parallel plate capacitor (Fig. 6.14(c)). A voltage develops between the electrodes in response to temperature change.

The detector output depends on the amount of radiation hitting the pyroelectric material. The transducer converts incident radiation into heat, thereby raising the temperature. This change in temperature gets converted to an electrical voltage signal that is then amplified through signal conditioning circuitry. Mostly, the pyroelectric transducers are used for radiation detection. The detection of intrusion is based on the human body's infrared emissivity. These transducers can also be used for ball bearing failure detection, intrusion detection, transient heating studies and gas flow monitoring.

6.11.1 Principle of Operation

To understand the principle of operation of pyroelectric effect, consider that a thin, parallel-sided sample of anisotropic solids material, such as a tourmaline crystal or barium titanate, cut in such a way that its crystallographic symmetry[1] axis is perpendicular to the flat surfaces. Bear in mind that the unit cells of pyroelectric materials have a dipole moment. Within the crystal the dipoles are so packed that the components of the dipole moment in each unit cell add up in the direction normal to the flat surfaces. The density of dipole moment of the material is known as the spontaneous polarization and is nonzero. The spontaneous polarization exists in the absence of an applied electric field and is equivalent to a layer of bound charge on each flat surface of the solid sample. The electrons or ions get attracted if found at the vicinity of the flat surface. If any conducting electrodes are connected to the surfaces via an ammeter, while changes in temperature exist, a current called pyroelectric current, will flow through the circuit. An increase in temperature causes the net dipole moment and, consequently, the spontaneous polarization to decrease. The quantity of bound charge actually decreases, and the redistribution of free charges to compensate for the change in bound charge results in a current flow. If the temperature of the sample is, however, maintained at a constant level the spontaneous polarization will remain constant and no current would flow.

The sensitivity of the detector increases if the temperature change is optimized. High sensitivity is achieved by considering the following.

- Materials having large value of pyroelectric coefficient.
- Materials should have small dielectric constant and specific heat.
- Design parameters with small thermal capacitance and conductance.

Ferroelectric $PbTiO_3$ (PTO) and $Pb_{1-x}Ca_xTiO_3$ (PCT_x) are very appropriate materials in this respect. The Curie temperature (CT) of $Pb_{1-x}Ca_xTiO_3$ (PCT_x) is primarily composition dependent. Its CT can be varied from 490–0 °C when x increases from 0 to 0.5. Therefore PCT_x is called a controlled material. The controlled material offers the benefit of choosing the operating temperature of the device. For instance, by properly deciding the value of x one can design a film with a Curie temperature slightly above the room temperature, thereby placing the maximum in the temperature dependent value of

[1] Symmetry plays an important role in crystallography. The ways in which atoms and molecules are arranged within a unit cells and unit cells repeat within a crystal are governed by symmetry rules (Robert Von Dreele).

pyroelectric coefficient at the desired device operating temperature. Curie temperatures of some materials are given in Table 8.1.

Bulk micromachining process is employed for the fabrication of the pyroelectric device (Fig. 6.14). The process is as follows.

- A p+ etch-stop epitaxial layer of 4 μm thickness is deposited on an n-type substrate.
- Suspended membrane is formed by anisotropic etching in EDP (Ethylenediamine Pyrocatechol)
- Deposition of PTO or PCT_x is performed by spin coating.
- Deposition of gold film is made to form one of the pyroelectric charge collection electrodes of the sensor. Counter electrode is formed by RIE (Reactive Ion Etching). Selective p+ layer provides this part.
- Lastly IR (Infrared) absorbing layers such as metal black and thin gold film are developed.

6.12 SHAPE MEMORY ALLOYS (SMA)

There exist some alloys which when heated transform from martensite to austenite phase. Martensite exists at lower temperatures, and austenite exists at higher temperatures. These alloys are called Shape

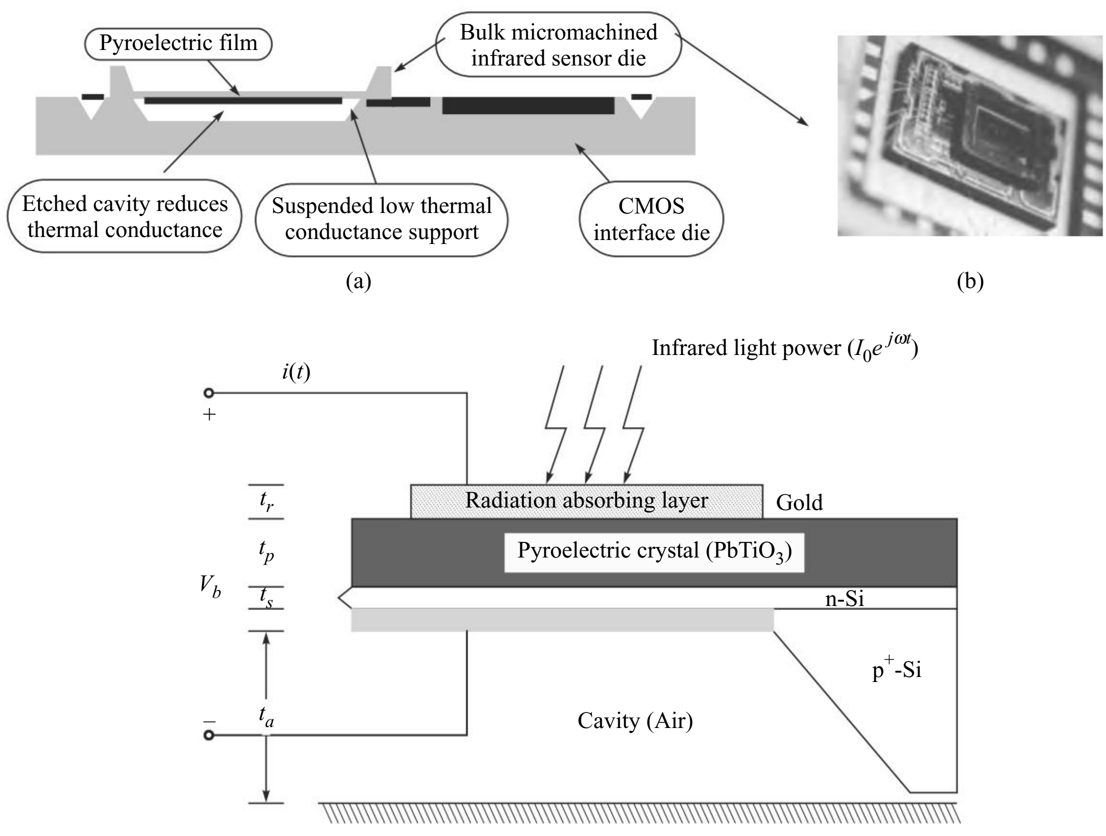


Fig. 6.14 *Pyroelectric sensor, (a) Structure, (b) The photograph (Source:* Chang et al, U. of California*), (c) Details of the structure (Source:* Ho et al., *IEEE T on Electronic Devices,* 46/12, 1999*)*

Memory Alloys (SMA). When an SMA is in martensite phase at lower temperatures, the alloy can easily be deformed into any shape. In the austenite phase, the alloy remembers the shape it had before it was deformed.

SMAs are active materials, which have the ability to return to a predetermined shape when heated. The SMA can be used for the design of actuator if operated above the memory transformation temperature (MTT), a temperature below which it has very low yield strength and is susceptible to deformation. MTT is the temperature that the alloy comes back to the original shape that it was before deformation. This further clarifies that when the material is heated above the MTT, it undergoes a change in crystal structure, which causes it to return to its original shape. This phenomenon provides a unique mechanism for actuation. Figure 6.15(a) illustrates the phenomena.

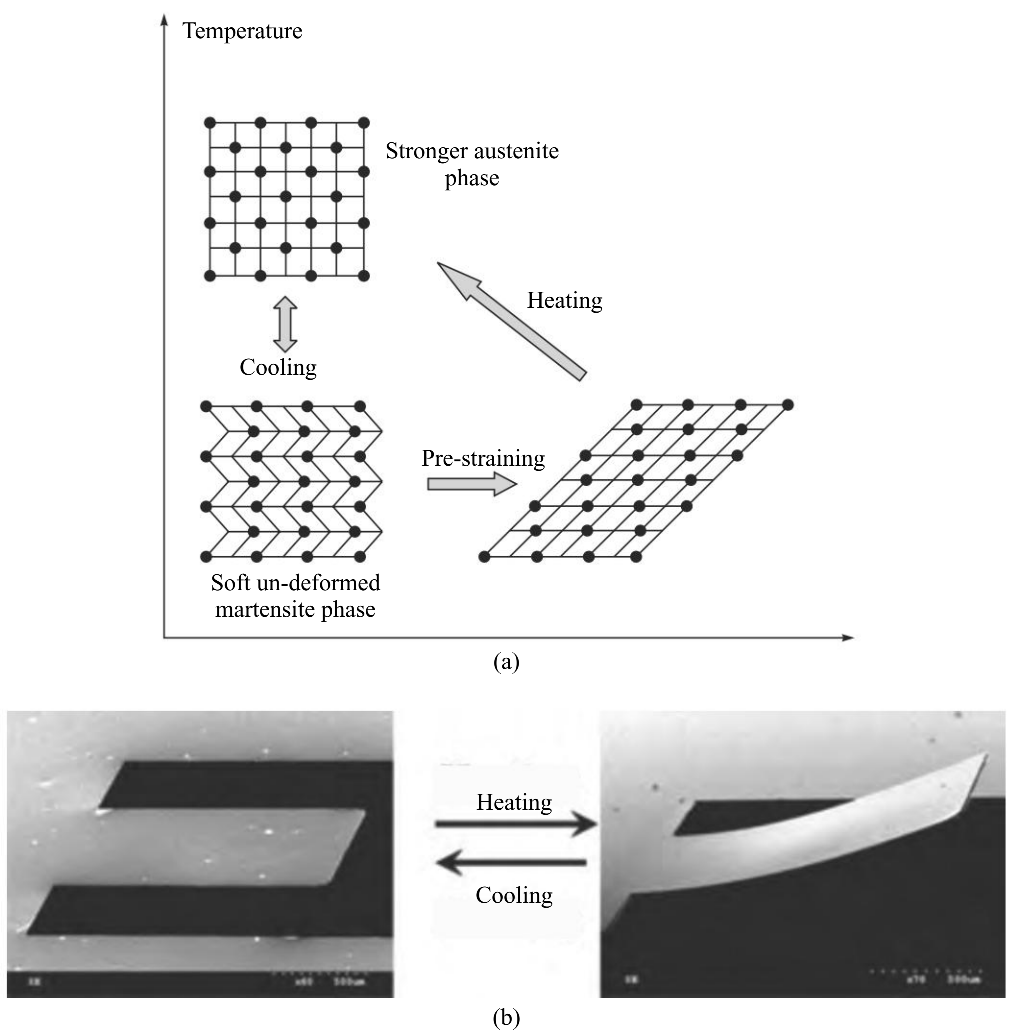


Fig. 6.15 *(a) Phase transformation phenomena of SMA (Courtesy:* Oulu University Library*), (b) TiNi cantilever showing the actuation during heating and cooling (Courtesy:* Fu, Luo, Flewitt and Milne: University of Cambridge, UK and Nanyang Technological University, Singapore*)*

The commonly used shape memory material is an alloy of nickel and titanium, called Nitinol. Nitinol is well known for its thermomechanical properties of superelasticity and shape memory effect. The alloy has very good electrical and mechanical properties with long fatigue life and high corrosion resistance. With excellent biocompatibility nitinol has emerged as a unique alloy. It is finding an increasing number of medical applications such as orthodontic wires, orthopaedic devices, guide wires, filters and components in minimally invasive surgical devices. Composition and subsequent processing characterize the properties and MTT point of the Nitinol. For instance, the transformation temperatures are extremely sensitive to a small variation in the Ni or Ti concentration. The sensitivity increases with Ni content in the alloy. Transfer temperatures can be altered by slight changes in composition, and by slight changes in heat treatment. MTT can be made precise, within 1 or 2 degrees of the desired temperature. Heating is the only way that most memory metals retain their original shape. Figure 6.15(b) shows SEM photograph of an SMA based actuator.

6.12.1 SMA Driven Micro-ball-lens Optical Switch

NASA's Jet Propulsion Laboratory has developed a Micro-ball-lens optical switch which is driven by SMA actuator. This is a prototype of low-loss mass-producible optical switch. Figure 6.16 is a simplified cross section of a microscopic optical switch. The figure is self-explanatory. The light is coupled from an input optical fiber to one of two side-by-side output optical fibers. The optical connection between the input and the selected output fiber is made via a microscopic ball lens.

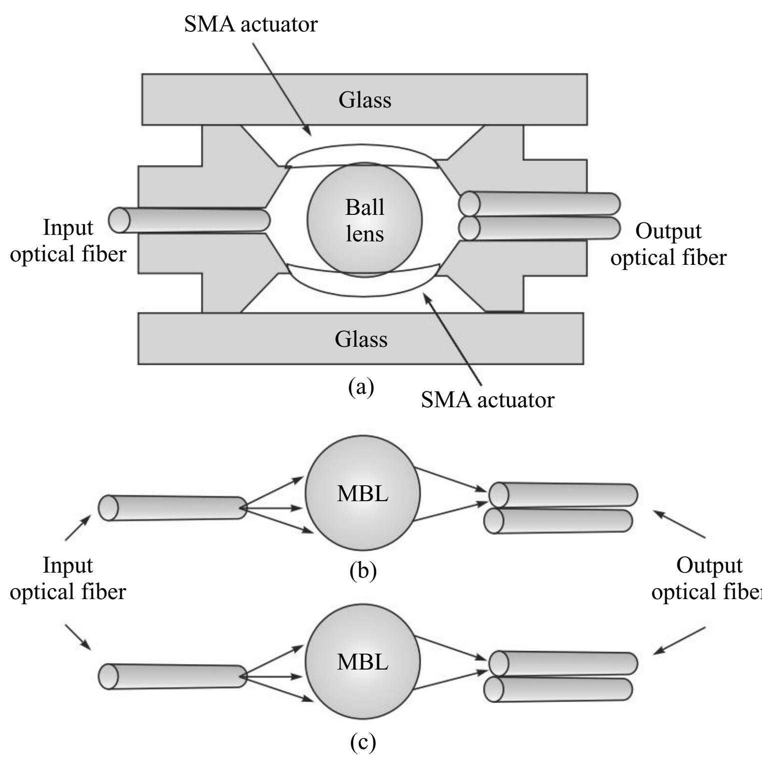


Fig. 6.16 *(a) SMA driven micro-ball-lens optical switch, (b) Micro Ball Lens (MBL) in upper position, (c) MBL in lower position, (Courtesy:* Eui-Hyeok Yang of Caltech for NASA's Jet Propulsion Laboratory*)*

Switching of the optical connections from one output fiber to another is performed by using a pair of thin film SMA actuators as shown in the figure. The actuators can toggle the lens between two resting switch positions.

The diameter of the lens is about 1 μm. The two actuators hold it. Although, the figure looks simple, the real design is very complex, partly made of netlike shapes. The optical fibers are held in V-grooves made up of silicon frame. This is a common technique. To effect switching, electric current is passed through one of the SMA actuators in order to heat it above its transition temperature, thereby causing the SMA to deform to a different 'remembered' shape. Once switching has taken place and the electrical current is turned off, the two SMA actuators are mechanically stiff enough so that the lens remain latched in the most recently selected position. Displacement of about 100 micron is the desirable range and is easily achieved. Further, the required switching frequency is 10–15 Hz. Typical insertion loss in such a device is about 2 dB, which is extremely low and considered as very good design characteristics.

6.13 U-SHAPED HORIZONTAL AND VERTICAL ELECTROTHERMAL ACTUATOR

There are two varieties of electrothermal actuators, namely horizontal and vertical. Both are U-shape in structure. A schematic diagram of a horizontal thermal actuator is shown in Fig. 6.17(a). In this design there is one hot arm. Figure 6.17(b) shows another design, in which there are two hot arms. The latter one is better in terms of efficiency. The schematic diagram of the vertical thermal actuator is shown in Fig. 6.17(c). Figure 6.17(d) shows another design of vertical type. The second design is better compared to the former type (i.e. Fig. 6.17(c)) because it can actuate in both directions: upward and downward. In all the designs distinct cold and hot arms are present. The cold and hot arms are made of the same material.

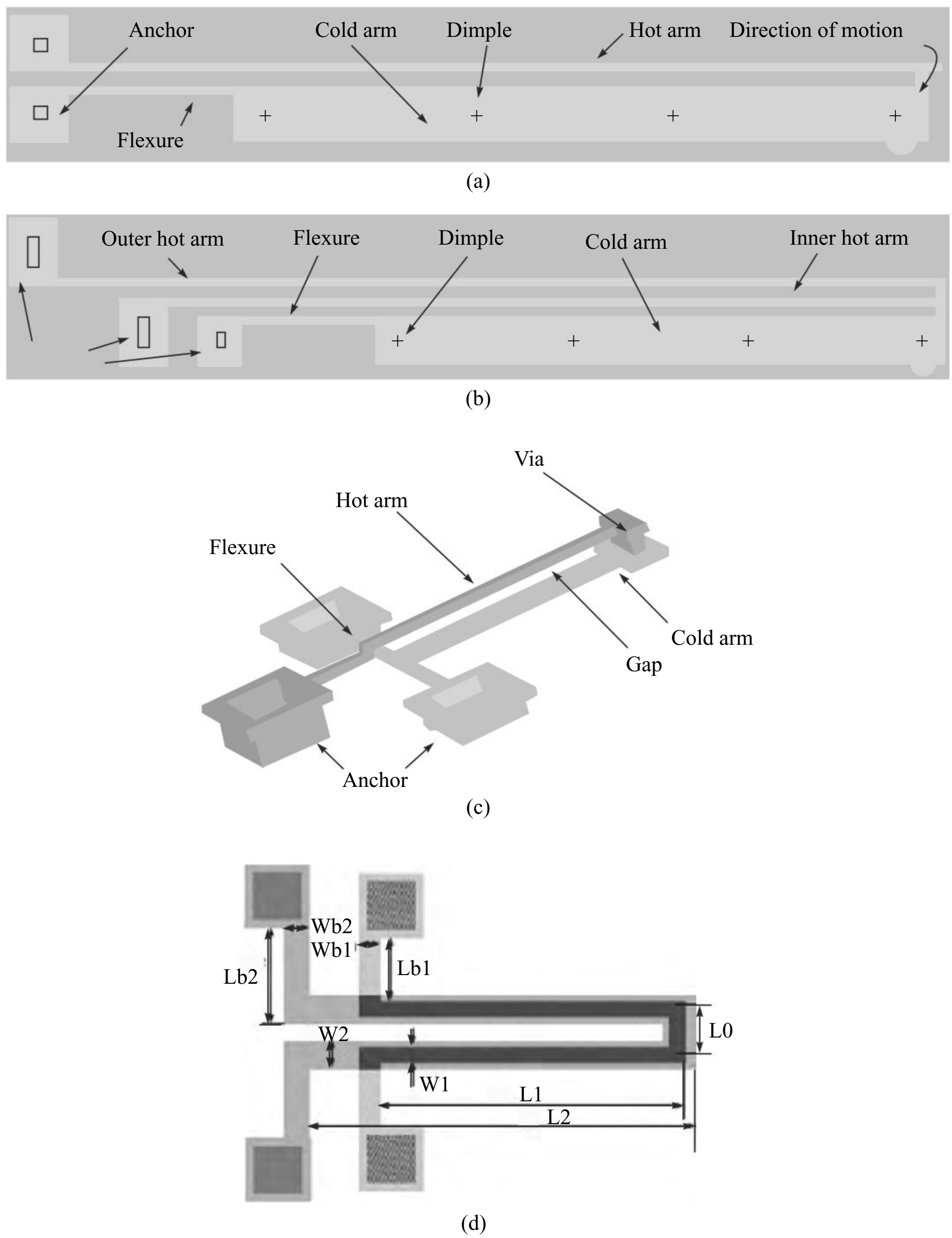
6.13.1 Horizontal Actuator

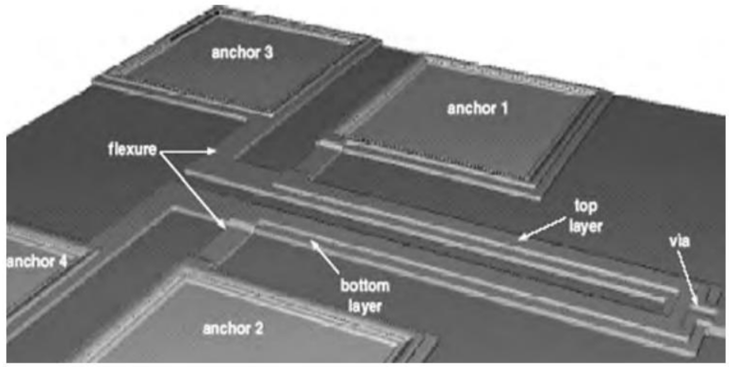
The hot arm is thinner compared to the cold arm. Note that the thinner the dimension, the larger is the electrical resistance. In order to achieve actuation, current is passed through both the arms. When current passes through one-hot-arm type actuator, the hot arm is heated to a higher temperature as it has higher resistance value. The actuation is ensured only because of asymmetric heating of the cold and hot arms. Therefore, the hot arm expands in length compared to the cold arm. The differential expansion makes it possible to deflect. The deflection appears as a rotation about the flexure. Flexures are natural candidates in MEMS construction. Behaving as links, there are two important features of these candidates: they allow movement such as joints, and sometimes transmit force or displacements. In case of two-hot-arm actuator the current is passed through the outer and inner hot arms. In this construction the flexure is always thinner than the hot arms. The actuation efficiency is higher as compared to one-hot arm based design since all the power consumed contributes only to the deflection of the actuator.

6.13.2 Vertical Actuator

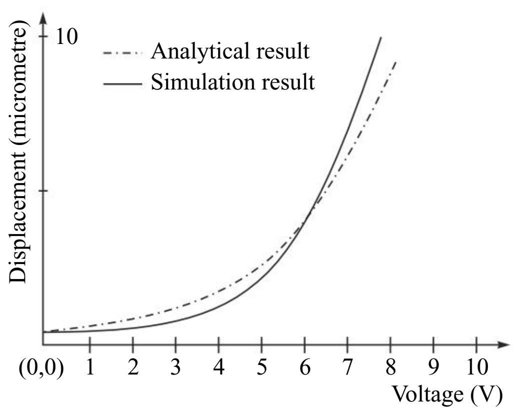
In case of vertical type actuator the structure of the hot and cold arms are such that the combined effect creates a vertical motion perpendicular to the substrate. By using a design shown in Fig. 6.17(c), the actuation can be achieved in one direction, whereas by using a design of type shown in Fig. 6.17(d),

bi-directional actuation can be achieved. Therefore, the latter type is sometimes referred to as U-shaped Bidirectional Vertical Thermal Actuator (BVTA).

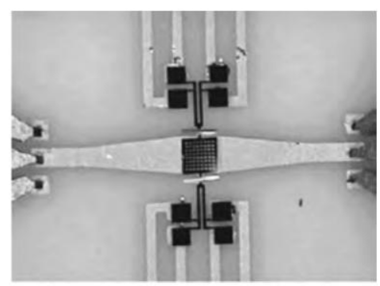




(e)



(f)



(g)

Fig. 6.17 *U-shaped horizontal and vertical electrothermal actuator, (a) One-hot-arm horizontal thermal actuator, (b) Two-hot-arm horizontal thermal actuator, (c) Vertical electrothermal actuator, (d) U-shaped vertical actuator (2D view), (e) Its 3D view, (f) Deflection of the tip of the BVTA w.r.t. input voltage, A tunable capacitor with two Bidirectional Vertical Thermal Actuator (BVTA).*
Source: Yan, J. *of J. Micromechanics and Microengineering,* 13 (2003), 14 (3004)

The one directional actuator has the hot arm, which is placed above the cold arm. The two arms are separated by the air gap. The arms are connected together at one end. The connections are made by, what is known as *via*. At the other end the arms are separately anchored on the substrate. Both the arms are joined at this end through a flexure. When the driving current passes through both the arms, thermal energy is generated. Actuation is observed as explained in the horizontal case. The tip of the actuator will move downward towards the substrate.

In case of BVTA, however, the actuation is achieved both upward and downward. The bending direction of the BVTA depends on the point of application of voltage signal. If the voltage signal is applied across the anchors 1 and 2 (Fig. 6.17(e)) the current will pass through the top layer and the actuation will be downward to the substrate. On the other hand if the voltage signal is applied across the anchors 3 and 4 the current will pass through the bottom layer and the actuation will be upward to the substrate.

6.14 THERMALLY ACTIVATED MEMS RELAY

The primary function of the relay is to transfer power. The basic action of relays is very simple, i.e. switching ON and OFF (Refer to Fig. 6.18(a)). The relays can either make or break the circuit. They offer low and high resistances, while ON and OFF, respectively.

Mechanical relay driven by magnetic coils are widely used in various places like control applications and in power sectors. Semiconductor relays are the basic elements in designing PLC (Programmable Logic Controllers). Refer the book *Mechatronics: Principles concepts and application* by McGraw-Hill for more details (Chapter 8). This section describes the basic principle of MEMS relays.

MEMS based relays are a recent development. As obvious, MEMS relays are miniaturized mechanical relays, which are fabricated based on MEMS technology. The constructional feature varies depending upon the application domain. While the basic action remains same but their design, fabrication process, manufacturing, and application significantly vary. In power sector, particularly, the applications have ranged from signal switching to power switching. The constructional requirement for signal switching and power switching are quite different. It should be pointed out that the signal switching operation handles less current than the power switching operation. For power switching applications, the required ON-state resistance is approximately 15 mΩ and the current-carrying capacity is several amperes, unlike signal switching case. Regardless of their applications, the desirable characteristics of the MEMS relays are:

- Good ON-state and OFF-state characteristics
- Large stand-off voltage when the relay is at the OFF-state
- Gap between the contacts should be minimum, usually 20–25 μm
- Low contact resistance. The contact resistance is a function of contact force. Therefore, the contact force should be consistent through its operational period.
- The switching speed should be very high. Expected typical switching speed is less than 3 ms.

Like other devices, the MEMS relay design considerations include formal specification of functional requirements as illustrated in Fig. 6.18(b). The functional requirement of a device hypothesizes aspects of structural configuration and actuation principles to be adopted along with favorable fabrication methods so that the final design will be able to optimally facilitate the desired functions.

Typical MEMS relays for power applications entails bistable double beam structure (Fig. 6.18(a)). Bistable structure has two stable states, and can be compared with the bistable multivibrator in electronic circuits. Thermal actuation is employed in the relay for achieving switching operation. MEMS relays can be fabricated through both bulk and surface micromachining process.

A schematic diagram of the relay utilizing the principle of thermal actuation is expressed in Fig. 6.18(c). At the center of the relay there is a bistable pre-curved double beam that moves laterally in the wafer plane. A mechanism called crossbar is attached to the double beam through a cantilever beam. When the double beam is in the position shown as solid lines the two contacts below the crossbar are not connected, and therefore the relay is now at OFF state. When the double beam snaps toward its second stable position, shown as dashed lines, it thrust the crossbar against the two contacts, and thus the relay is now at ON state. The relay has two thermal actuators. Each actuator has two beams. Thermal actuator-1 deflects the bistable double beam downward to close the relay, whereas the thermal actuator-2 deflects upward to open the relay. Figure 6.18(d) shows its MEMS photograph.

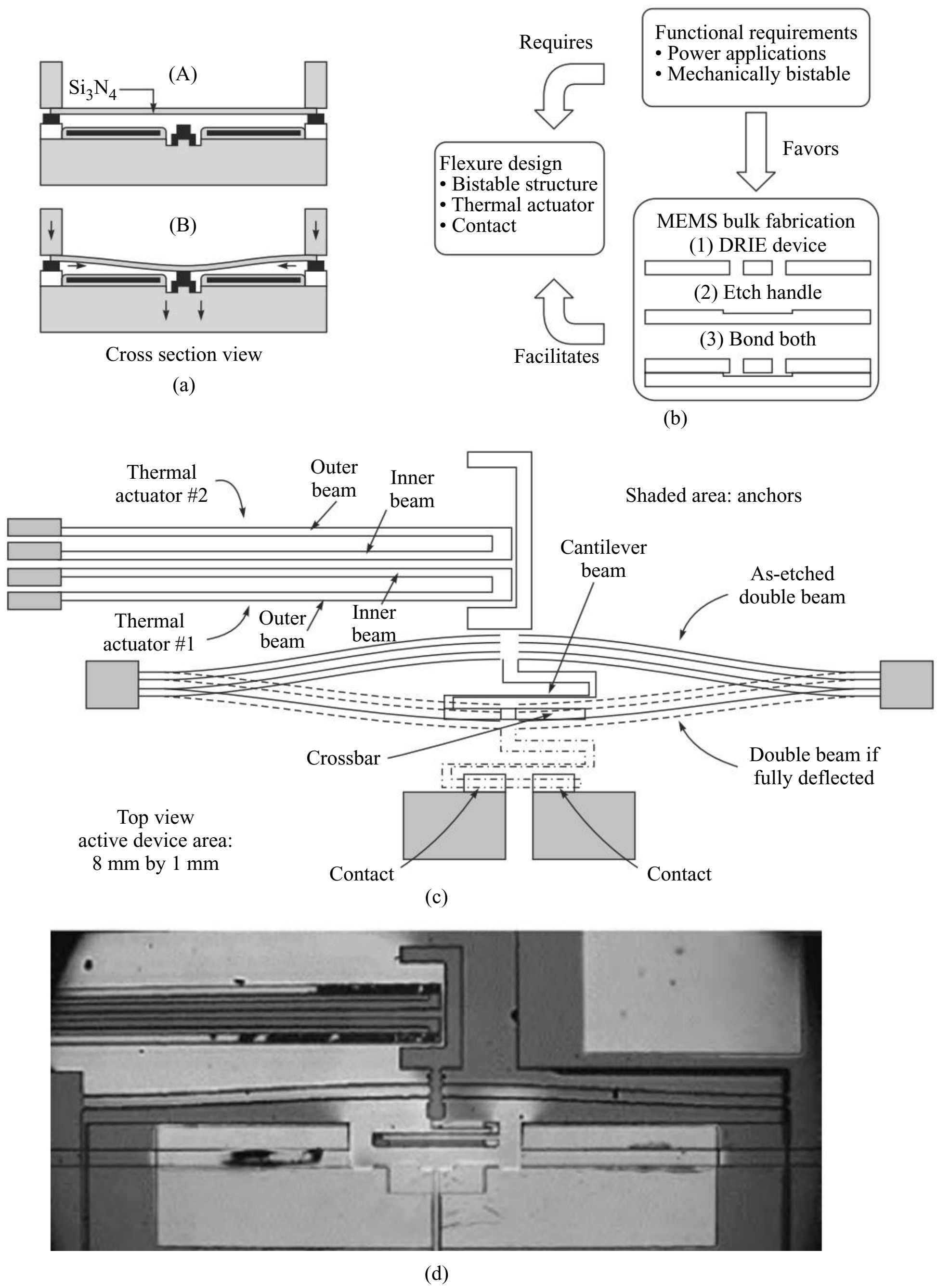


Fig. 6.18 *Thermally activated MEMS relay, (a) The requirement specification, (b) A simple switching illustration of a bistable structure relay, (c) Illustration showing thermally activated MEMS relay, (d) Micromachined thermal MEMS relay*

Source: Jin Qiu, *Dissertation*, MIT, 2003

The thermal actuation requires either one of the following combinations. The two beams of the actuator are made up of,

- Same material with different dimensions
- Same dimension with different materials (having different coefficient of expansions)

When the beams are heated by employing either of the above configuration, one of the beam will elongate with respect to the other, resulting a deflection and hence the actuation. The beams can be heated by passing the electrical current through them.

6.15 MICROSPRING THERMAL ACTUATOR

Electrothermal actuators, called heatuator, deliver large forces with large range of operation. The heatuator utilizes different structural configuration to enhance the displacement range. The structures are made from the same material to induce varying in-plane lateral expansions and hence motion. The displacement range can be significantly improved by designing spring type actuating element. The spring type actuating element based thermal actuators are referred to as chevron[2] actuator. The actuator consists of a number of chevron structures. Figure 6.19(a) shows a planar spring-like chevron structure. Note that the chevron structure uses two types of materials. Figure 6.19(b) shows an SEM photograph of a freestanding microstructure with 8-chevron sections in series.

The v-like structure (spring) are made from the material having high thermal coefficient of expansion, while the straight insulator bars that forms the cross-linkage are made from the material having low values of thermal coefficient of expansion. This is so because the linkage bars constrain the displacement in the x-direction. When the actuator is heated up through electrothermal method, the thermal expansion of the active material will make it possible to have a displacement in the y-direction

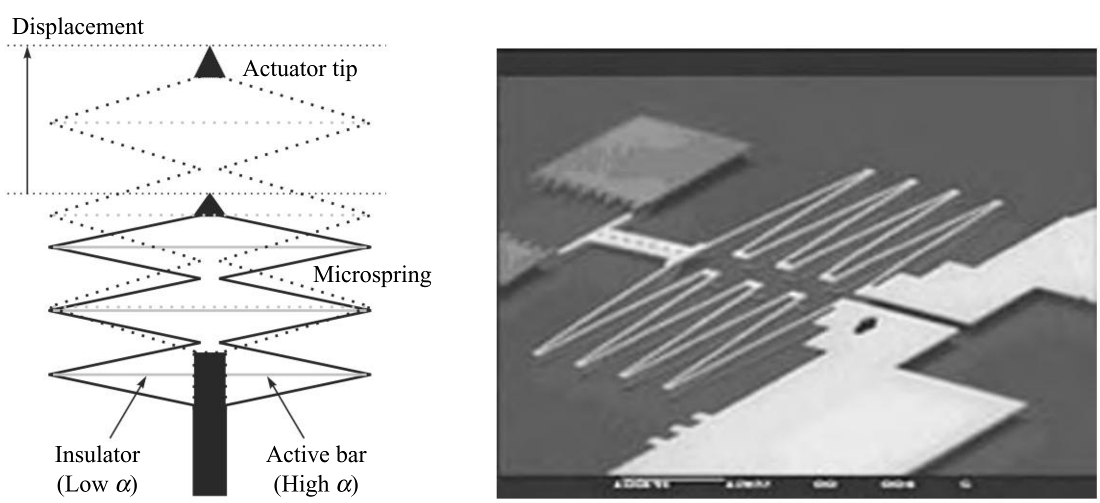


Fig. 6.19 *(a) Schematic drawings of a chevron structure: three-ring spring, (b) A freestanding microstructure with 8-chevron sections in series*
Source: J. Micromech. Microeng. 15 (2005) 1527–1535

[2] Chevron: A regular zigzag pattern of straight line generally disposed horizontally.

only. By connecting a number of chevron sections in series, it is possible to produce a large displacement up to 100 μm. The lengths of the active bars are about 200 μm. Electroplated Ni and SiO_2 (also Si_3N_4) can be used as active and insulating material for the actuator, respectively.

6.16 DATA STORAGE CANTILEVER

One exclusive example of thermal MEMS design is the probe-based cantilever structure for high-density thermo-mechanical data storage employing atomic force microscopy (AFM) method. MEMS-based scanning-probe data storage devices are emerging as potential ultra-high-density, low-access-time, and low-power alternatives to conventional data storage device. The implementation of probe-based storage system (Fig. 6.20) uses thermo-mechanical means, as described below, to store and retrieve information in thin films. The device includes thermal position sensor that provide position information to the servo controller.

The system is based on design of microcantilever tip that exerts a constant force on a polycarbonate sample (disk) and induces localized softening and deformation during heating phase. In an AFM-based thermo-mechanical data storage system digital information is represented as submicron pits (groove) on a rotating disk. The sharp tip of a cantilever is kept in continuous contact with a spinning polycarbonate disk by a weak loading force on the order of 10^{-7} Newton. For writing operation, the cantilever tip is heated above the glass transition temperature (GTT) of polycarbonate. GTT is the critical temperature at which a polymer (i.e. polycarbonate thin film on the disk) loses the properties of glass and obtains the properties of an elastomer. Elastomer is another property which defines that a polymeric material experiences large and reversible elastic deformations. The GTT of the polycarbonate is approximately 120 ^{o}C. While heating, the cantilever is also subjected to the loading force (10^{-7} N) thereby melting data pits onto the polycarbonate disk. The heat is generated by a bias current along the cantilever. The *thermal time constant* (TTC) for cooling of the heated cantilever tip governs the rate at which it can achieve writing. The required TTC is about one second. For achieving reading operation the cantilever deflection due to tip rides over the pits has to be measured. The precise measurement is achieved by the use of a separate cantilever integrated with a piezoresistive displacement sensor followed by measuring circuitry.

One of the challenges in building such devices is the accuracy and the latency required in the navigation of the probes over the polymer medium. The design and characterization of a servomechanism to achieve precise positioning in a probe-based storage prototype is extremely critical.

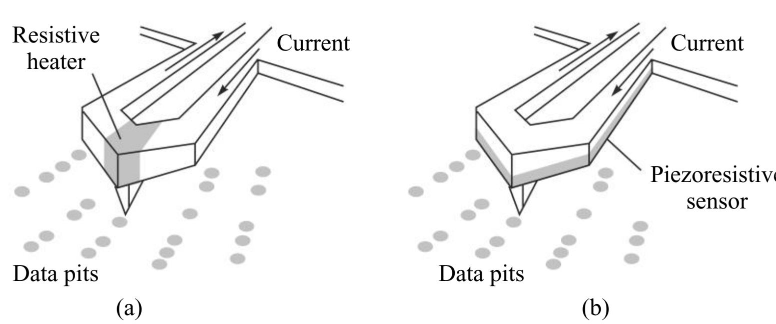


Fig. 6.20 *Principle of thermo-mechanical data storage using (a) cantilevers with built-in heaters for writing and (b) integrated piezoresistive sensors for readout*
Source: Chui et al., J. of MEMS, 7/1, 1998

The microcantilever tip is very sharp. Typically the radius of curvature of the cantilever tip is about 500 Å. At this dimension it is possible to achieve a data storage density of 20–50 Gb/in^2. For reading operation, the system is capable of measuring the rides of about 10 Å. In order to avoid wear of the cantilever tip as well as the polycarbonate disk the integrated system should be able to deal with low amount of loading force which is usually kept below 10^{-7} N. For this it should have stiffness value typically below 1 N/m. Further, the resonant frequency should be as high as possible. Moreover, it is very important to consider the mass of the cantilever. Bit densities up to 30 Gb/in^2 (50 times CD-ROM) and reading rates up to 1.2 Mb/s have already been reported.

6.17 SUMMARY

MEMS devices refer to integration of mechanical and electronic components on the micron size and include 3D lithographic features of various geometries. They are manufactured using planar processing much similar to IC processes but advanced method such as surface micromachining and/or bulk micromachining. MEMS groups of devices, which explore the thermal principle and effects, are called thermal MEMS. The thermal principles are thermoelectric effect, Peltier effect and thermoresistivity.

This chapter deals with thermal sensors and thermal actuators based on MEMS technology. The chapter describes the fundamental principles of thermal sensing and actuation. Basics of conduction, convection and radiation are explained. The feature and characteristics of thermistors and their governing equations are included. The construction and principle of operation of thermodevices such as thermodiode and thermotransistor are described. Thermocouples are the important thermal sensors. The principle of operation of thermocouple and thermopiles are presented. The heat pumps utilizing Peltier effects are being designed for a wide range of applications including computing. The constructional detail of a generalized Peltier heat pump and heat sink is described. Moreover, operational principle of hotwire and microhotplate based thermal flow sensor are explained. Microthermovessels play important role for DNA analysis. A brief introduction in this respect is considered.

Shape Memory Alloy (SMA) are the material whose shape change when subjected to heat. These materials can be used for microactuation applications. Further, the design and construction of U-shaped horizontal and vertical uni- and bidirectional thermal actuators are discussed comprehensively. Relays are important building blocks in micromachineries. Bistable MEMS relays for power system applications are explained. Micromachined thermocouple probe based feedback system can be used for data storing applications. An example in this respect is given towards the end of the chapter.

Points to Remember

- Thermal MEMS refers to the microsystems whose functionality relies on heat transfer.
- Thermal MEMS are categorically divided into three types: thermal sensor, thermal actuator and data storage device.
- Like all other sensors, the thermal sensors convert physical phenomenon through a fundamental process called transduction.
- The transduction phenomenon is observed in three ways, namely thermoelectric effect, thermoresistive effect and the pyroelectric effect.
- Thermoelectric effect implies that a voltage is developed in a loop containing two dissimilar metals when the end junction of the loop is heated.

- The opposite of the thermoelectric effect is called Peltier effect, in which current flow causes a temperature difference between the junctions of different metals.
- Thermally activated important actuation mechanisms are pumping and valving, which are usually employed in the bio-analytical microsystems such as *lab-on-a-chip* (loac) devices.
- Thermal actuation is based on electrothermal energy density transformation which is given by $E = V^2/\rho L^2$.
- Energy is always present but not visible.
- Thermal energy is quantified by temperature. There are several ways to express temperature and heat. Temperature is considered as the physical property of a material that signifies whether it is hot or cold. Two elements that are at different temperatures transfer thermal energy from hotter region to colder region.
- The faster moving molecules collide with each other more frequently and as a result of which they need more space, apparently decreasing the density of the molecules.
- Thermal energy can be transferred in three ways, namely by conduction, convection and radiation.
- Conduction is a process that transfers energy from one molecule to another.
- When a material is heated the flow of heat makes it possible to increase the kinetic energy of the atoms subsequently causing more distance between the atoms. With regard to such expansion materials are classified whether they are isotropic or anisotropic type. In an isotropic material, the expansion occurs uniformly in all dimensions, whereas non-uniform expansion occurs in case of anisotropic materials.
- The heat capacity of a material is the amount of heat required to change its temperature by one degree.
- Convection is another process in which the movement of liquid such as water or gas causes heat transfer.
- A thermistor is a piece of sintered semiconductor material, which exhibits a relatively large change in resistance proportional to a change in temperature. Thermistor possesses negative temperature coefficient of resistance, i.e. the resistance of the piece decreases with increase in temperature.
- Thermistors can also be designed from metals. Metal thermistors have positive temperature coefficients of resistance.
- Thermistors are manufactured by *sintering* process, which consists of (i) the preparation of the metal oxide, (ii) milling and blending, (iii) heat-treatment, (iv) addition of electrical contacts, (v) assembly into a device, and (vi) protective coating.
- Semiconductor diodes and transistors specially designed to meet the requirements of the temperature measurement are individually called *thermodiodes* and *thermotransistors* respectively and commonly called *thermodevices*.
- Thermotransistors are more sensitive as compared to thermodiodes.
- A thermocouple is a transducer that consists of two dissimilar metals, joined together at one end, called thermocouple junction. When the temperature of this junction is different from the temperature of other parts an EMF (Electromotive Force) is generated. The EMF produced is known as *Seebeck* EMF or simply *thermovoltage.*
- Thermal probes are devices specifically designed for precision monitoring of temperature in a variety of applications, including photothermal absorption, spectroscopy, subsurface imaging, photolithography research, micro-calorimetric analysis, scanning microscopy.
- The Peltier effect is useful in manufacturing the devices for heat pumping applications. Such a pumping device is called heat pump or Peltier Module (PM).
- When current is passed through the module one side becomes cold and other side becomes hot. The effect is reversible by passing the current in the other direction. Such a configuration called a heat pump because the heat is given out or pumped out of the cold surface and is deposited on other surface.
- The MEMS flow sensor can also be used in a pulsating flow and in a vibrating environment.
- In thermal flow sensor, the gas to be measured is allowed to flow across the surface of the sensor. There is a heater that heats the surface. The passage of the gas cools the sensor surface and the resulting surface temperature is measured.
- Thermal isolation improves measurement sensitivity and efficiency.

- DNA replication means the use of existing DNA as a pattern or model for the analysis and synthesis of new DNA strands.
- PCR requires accurate cycling of the liquid sample temperature with operational range between 55 and 95 degree Celsius.
- Pyroelectricity is a temperature dependent electrical polarization phenomenon.
- Mostly, pyroelectric transducers are used for radiation detection.
- SMAs are active materials, which have the ability to return to a predetermined shape when heated.
- The commonly used shape memory materials are an alloy of nickel and titanium, called Nitinol.
- There are two varieties of electrothermal actuators, namely horizontal and vertical. Both are U-shaped in structure.
- The primary function of the relay is to transfer power. MEMS relays are miniaturized mechanical relays, which are fabricated based on MEMS technology.
- One exclusive example of thermal MEMS design is the probe-based cantilever structure for high-density thermo-mechanical data storage employing atomic force microscopy (AFM) method.
- In an AFM-based thermo-mechanical data storage system digital information is represented as sub-micron pits (groove) on a rotating disk.
- For writing operation, the cantilever tip is heated above the glass transition temperature (GTT) of polycarbonate. GTT is the critical temperature at which a polymer (i.e. polycarbonate thin film on the disk) loses the properties of glass and obtains the properties of an elastomer.
- For achieving reading operation the cantilever deflection due to tip rides over the pits has to be measured. The precise measurement is achieved by the use of a separate cantilever integrated with a piezoresistive displacement sensor followed by measuring circuitry.
- The microcantilever tip is very sharp. Typically the radius of curvature of the cantilever tip is about 500 °A. At this dimension it is possible to achieve a data storage density of 20−50 Gb/in^2.

Exercises

1. Give a precise definition of thermal MEMS.
2. Mention three important thermal principles based primarily upon which the thermal MEMS work.
3. Elucidate the basics of heat transfer process.
4. Write short notes on the following.
   (a) Thermistors
   (b) Thermodevices
   (c) Thermocouple and thermopile
5. Draw the schematic diagram of a thermocouple probe and state how the probe tip can be used for 3D topographic applications. Draw the equivalent circuit of the thermocouple probe and define the terms. List the primary process sequences of fabricating a typical thermocouple probe.
6. Discuss the design and principle of Peltier effect heat pump. Give some applications of heat pumps.
7. Describe the principle of operations of various types of MEMS thermal flow sensors. Draw the schematic diagram of a differential type thermal flow sensor and explain. What do you mean by thermal isolation and how is it achieved?
8. Write short notes on microhotplate type gas sensor.

9. What is the function of MEMS thermovessel chip? What are the basic design requirements of most of microthermovessels?
10. Define pyroelectricity and Curie temperature (CT). Write down the composition of PTO. How can CT be modified? Why is the modification of CT necessary? Write down the fabrication steps employed in designing the pyroelectric MEMS sensor.
11. What is SMA? How does the SMA react to heat? How can SMA be used as a microactuator? Can the alloy be used as a sensory device?
12. Describe in detail the design, construction and principle of operation of various types of U-shaped horizontal and vertical electrothermal MEMS actuator. State the relative merits and demerits if any.
13. What is a relay? Draw the schematic diagram of a thermally activated MEMS relay. How do thermally activated MEMS relays work? What desirable characteristics should an MEMS relay have?
14. Write a short note on data storage cantilever.

Chapter 7

Micro-opto-electromechanical Systems

Objectives

The objective of this chapter is to study the following.

- Properties of light and their exploitation with respect to MOEMS
- Applications of MOEMS
- Fundamentals of MOEMS
- Optical switching (principle, concept, design, and applications)
- Spatial Light Modulator (SLM)
- Introduction to Digital Micromirror Device (DMD™)
- Introduction to Grating Light Valve (GLV™)
- Beam splitters and microlenses
- Micro Optic Waveguide (MOW) based tuning
- Stress/Strain measurement

7.1 INTRODUCTION

Equipment or a system is built by using discrete components and elements. The discrete components and elements are called *singular* devices. Many examples can be given. Electronic circuits and systems are designed by the use of singular electronic components and electrical elements such as transistors, diodes, resistors, capacitors, and so on. Once designed, the system handles electrical signals for its operation. In another example, an optical system such as an optical transceiver (A transceiver is a system that is capable of transmitting and receiving the optical signal from one unit to another unit) is designed by using singular optical devices and components such as light emitter, waveguide, directional coupler, and so on. Once designed, in this case however, the system handles optical signals for its operation. There are some other devices, which are not singular but *hybrid* in nature. These systems can handle many types of signals such as electrical, optical and RF. These hybrid devices are expressed in a specific manner.

7.1.1 Hybrid Systems

Opto-Electro-Mechanical (OEM) devices are claimed to be hybrid type devices. OEM devices utilize the principles, properties and methods of optical, electronics/electrical and mechanics in order to make themselves operational. When the OEM devices are designed at the microscale level they are called Micro-Opto-Electro-Mechanical devices or MOEM in short. Thus, combining optical theory with the microelectronics and micromechanics creates a new and broader class of microdevices popularly known as MOEMS (Micro-Opto-Electromechanical System) device. Figure 7.1 gives an idea as far as expression of MOEMS devices with regard to synergistic integration of diverse principles and methods are concerned. As can be seen, the MOEM systems are characterized by optics, electronics and mechanics. Because of involvement of optical theory and methods, MOEMS technology requires a different set of rules for operation as opposed to normal MEMS world. MOEMS has been considered as a new technology, which involves complexity. Its design method is not as simple as that of classical optoelectronic devices.

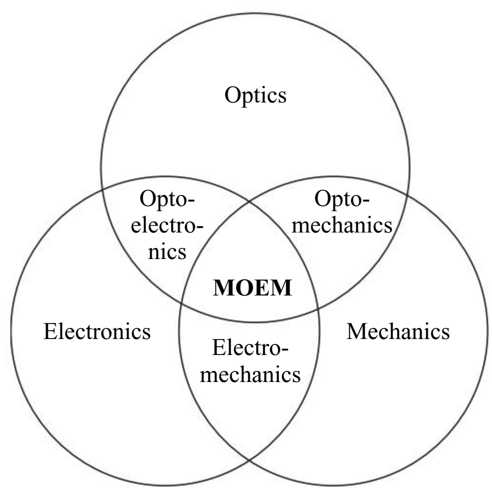


Fig. 7.1 *In illustration of the hybrid systems—MOEMS*

7.1.2 Applications

MOEMS have emerged to provide unparallel functionality in telecommunication applications. As a specific example, for instance, a traditional tunable micro-lasers and optical switching based on MOEMS technology significantly improves the capability of wavelength division multiplexing (WDM) systems. Broadly, MOEMS devices are found in beam splitting, beam combining, beam shaping, phase array and display applications. Manufacturers of MEMS devices are forecasting new opportunities of MOEMS outside of the telecommunication industries including: information technology, healthcare, industrial test and measurement, military and control sectors. Some applications of MOEMS are presented in Table 7.1. Typical systems-level applications are listed below.

- Arrays of micromirrors for digital image processing (projection, printing)
- Optical switches and routers, variable attenuators, shutters (optical communications)
- Tunable sources and filters, reflection modulators, spectral equalizers (optical communications, optical sensing, spectral analysis)

Table 7.1 *Applications of MOEMS*

Broader applications	*Specific applications*	*Comm^n applications*
Biomedical	Scanning	Switches
Automotive	Projection	Variable attenuators
Industrial maintenance & control	Display	Equalizers
Domestics	Printing	Modulators
Space and astronomy	Sensing	Dispersion compensators
Environmental monitoring	Data storage	Deformable micro mirrors
Consumer products	Micro motors	Optical chopper
Scientific equipments	Sensors	Digital or analog circuits
Microrobotics	Gyroscopes	Optical interconnect
Process control	Image transfer system	Micro-gratings

- Microscanners (image processing, bar code reading, obstacle detection)
- Deformable membranes for adaptive optics (astronomy, ophthalmology, FSO, defense)
- Free-space microoptics (diffractive and refractive microlenses)
- Guided optic devices
- Optical sources and photodetectors

7.1.3 Advantages

Although the ultimate speed of MOEMS is unlikely to compete with solid-state optoelectronic devices, but the precision that can be achieved with these systems is such that they contribute good performance because optical signaling is immune to noise and interference and bears negligible signal degradation feature. As a matter of fact the technology is being accepted as the preferred choice in the areas where the demand is for the implementation of flexible all-optical (AO) signal methods. The main advantages of MOEMS are,

- Miniaturisation
- Accuracy
- Insensitivity to electromagnetic interference
- Can be used in harsh environments
- Secured
- High sensitivity and selectivity
- Reliability and availability
- Mass production capability hence low manufacturing costs

Since bandwidth and power consumption are the key issues, using AO signal methods and advanced technologies for achieving better Quality of Service (QoS), MOEMS design need improved optical interface for time-efficient ways to make system integration.

7.2 FUNDAMENTAL PRINCIPLE OF MOEMS TECHNOLOGY

MOEMS technology enhances the level of perception and control. Directivity, sensing and actuation are fundamental to MOEMS. Evidently, the MOEMS technology is based on the movement of micro optical elements that in turn manipulate light in one-, two- or three-dimensional space. Movement of a micro optical element in conjunction with involved optical principle can make it possible to manipulate a light

beam. The parameters and principles are technologically combined to achieve complex operations on the light beam. The end effect is called *dynamic manipulation* of light. The *parameters* that are involved in the dynamic manipulation are amplitude, wavelength, temporal delay and spatial re-alignment. The involved optical *principles* are reflection or deflection, diffraction, refraction, fringing and polarization. The approach to carry out manipulation dynamically using miniaturized optical elements is considered as the fundamental principle of MOEMS technology. The fundamental principle is exploited in order to achieve a wide range of operations through application specific MOEMS devices.

Useful application specific MOEMS devices are microlens, micromirror, beam splitter, grating, waveguide and modulator. The lens is used for focusing purposes. Mirrors are used for achieving reflection whereas gratings are used for implementing diffraction, deflection and also spectral resolution. The optical beams can be added and separated by the use of beam splitters. The waveguide guides or carries the optical wave. The role of modulator is to change the intensity of the light. It can attenuate the strength of light. Broadly, MOEMS can be synthesized into two sub-systems; the micromechanics and microelectronics. The micromechanics aspect of the MOEM assures the spatial structure of the device so that 3D movement and actuation of physical parts will be possible. The microactuation could be linear or rotary motion. The microelectronics plays two roles; handling and processing of electrical signal and facilitating the availability of driving voltage or current for achieving linear or rotary motion of the physical parts mentioned above. Apparently, the MOEMS is a kind of product that has merged micro optics and MEM technologies to take advantage of the following factors.

- The technology is compatible with VLSI circuit processing, ensuring that the final device can be produced in volumes at low cost.
- Advanced microsystems can be tailored

7.3 REVIEW ON PROPERTIES OF LIGHT

Light is a complex signal that has long been explained with rays and wavefronts. Many scientists have already reported the dual behavior, i.e. particle-like and wave-like behavior of light. The particle-like behavior concludes that light is composed of particles. On the other hand, the wave-like behavior justifies continuous nature of light. These dual properties of light have been scientifically experimented in order to describe most of the useful characteristics such as reflection, refraction, diffraction, interference, polarization and Doppler's effect.

Reflection Reflection is defined in terms of bouncing of light. Reflection of light occurs when the light wave meets a surface or a second boundary that does not absorb the energy of the incident wave and bounces it from the surface.

Refraction The speed of light varies from medium to medium. If light passes from one transparent medium to another, its speed changes, and as a result it does not travel in the same direction as it was traveling in the previous medium. This in turn makes the light to bend while traveling through different mediums. The amount of bending depends on two factors, the (i) *refractive index* of the mediums and (ii) incident angle. Incident angle is the angle between the light ray and the line perpendicular to the surface separating the adjacent two mediums. The bending is measured in terms of angle of refraction that is the angle between the light ray in the second medium and the normal.

Diffraction As light is a wave, it has the capability to bend around corners of the obstacle. The fact behind this bending is that all points along a *wavefront* act as if they were point sources. If the wave

meets a barrier with a small aperture of the order that is comparable to the wavelength of light, the *effective light sources* are blocked except at the point of opening. The light coming through the opening now behaves as a single point source and has to radiate in all directions, instead of just passing straight through the aperture. This property of light is defined as diffraction. Diffraction is thus the slight bending of light as it passes around the edge of an object. The amount of bending is essentially related to the relative size of the wavelength of light to the size of the opening. That is if the opening is much larger than the wavelength of the incident light then the bending is almost unnoticeable. On the other hand, if they are comparable or equal, the amount of bending is significant and can easily be seen with the naked eye.

Interference Interference is the superposition or overlapping of two or more waves resulting in a new wave pattern. Interference causes cancellation or an amplification of the wave at that point. The principle of superposition theory applies to interference. This states that the resultant magnitude is equal to the sum of the magnitudes of different waves at that point. If a *crest* of a wave overlaps on a crest of another wave at the same point then the crests interfere *constructively* and the resultant wave amplitude will be greater. Conversely, if a crest of the wave overlaps a *trough* then they interfere *destructively*, and the overall amplitude will be decreased.

Polarization Polarization is a property of light. The light is considered as a mixture of *many* kinds of waves with *multitude of different wavelengths*. Each of these different waves vibrates in different directions as the light travels. If the number of waves vibrating in a particular direction is more as compared to other waves in the mixture, then the vibrating waves seem to overpower the other waves and the light is said to be polarized in that direction. Quantitatively, the polarization is expressed in terms of orientation of electric and magnetic field vectors of an electromagnetic wave (Light wave is an electromagnetic wave). It is possible to transform unpolarized light into polarized light. There are a variety of methods of polarizing the light. Some important methods are by transmission, reflection, refraction and scattering.

Doppler's effect Light from moving objects will appear to have different wavelengths. In fact the wavelength of light shifts when a moving object emits it. This effect is called the Doppler shift. The shifting of wavelength depends on the relative motion of the source and the observer. The amount of shift increases with the speed of the object. If the object is coming closer, the shifting is toward shorter wavelengths and is called *blue shifting.* If the object is going away the shifting is toward longer wavelengths and is called *red shifting.*

7.4 LIGHT MODULATORS

Modulation means "to change". Light modulator modulates or changes some of the parameters of the light. The important parameters are wavelength and direction. Modulators are often called manipulator. The modulators can also fragment, combine and polarize the incident light as and when required. All these manipulations work on various principles. Accordingly, there are various types of modulators. Two very important different technologies have been developed and studied extensively. The first one is based on micromechanics that involves movement of physical parts and utilizes reflection and diffraction. The two categories differ with regards to switching speeds color generation, efficiency and light source requirements. The second type of technology involves liquid crystals and uses the advantage of polarization. The polarization based modulation technology is at the research stage for which this chapter deals with the former types.

Modulation of light is essential in high-end display applications. The TV image and computer screen displays are called high-end display devices. The modulators used for these high-end applications are primarily of two types, such as square-structured micromirror 2D array type and ribbon-structured micromirror type. The square-structured MOEM modulators uses the binary (ON/OFF) switching method and the modulators are called Digital Micromirror Device™ (DMD™) which uses the reflection principle, and this technology has been defined what is known as Digital Light Processing™ (DLP™) technology. The ribbon-structured modulators are called Silicon Light Machines (SLM), which use diffraction principle, and this technology has been defined what is known as Grating Light Valve™ (GLV™) technology. Detailed explanation and working principle of these modulators are presented in the latter part of the chapter.

Other simple light modulators are beam splitter and microlenses, which are described in the following section. Light modulators are designed by the use of MEMS technology. Most of the MOEM devices are some sorts of light modulators.

7.5 BEAM SPLITTER

A beam splitter (BS) (Fig. 7.2) is a class of optical device that divides a light beam into two separate beams, hence the name beam splitter. Beam splitter can be a diffractive optic device. It is believed that the BS is an optical version of a copying machine. The splitted output beams can be sent in any direction of interest.

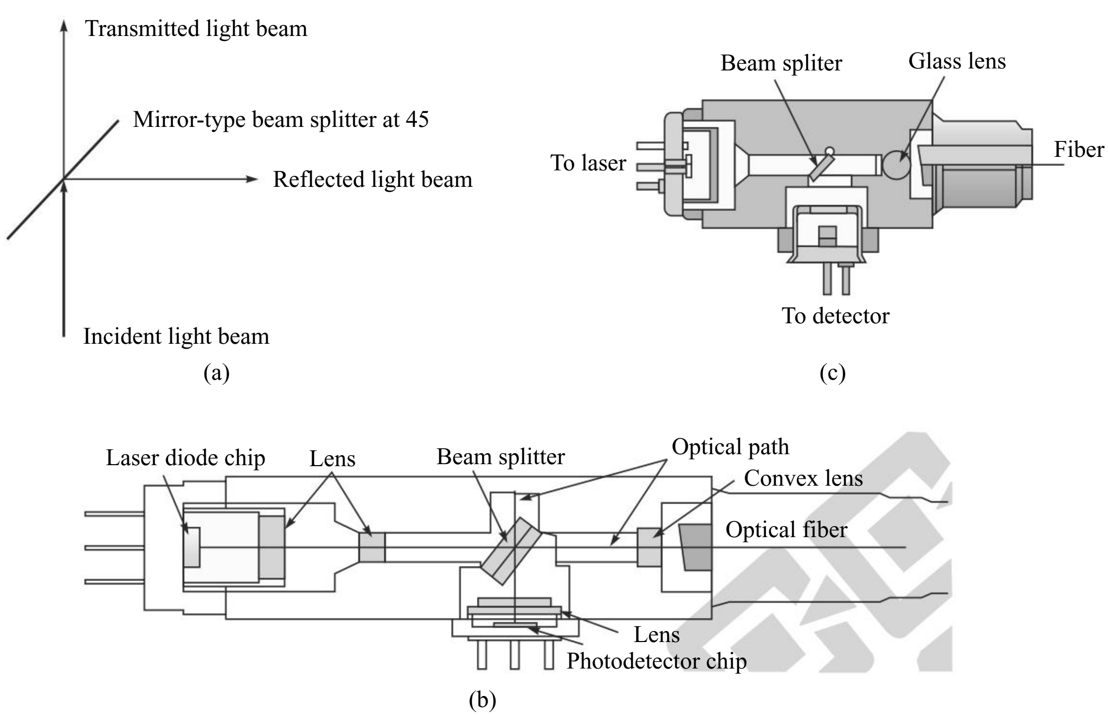


Fig. 7.2 *(a) Principle of beam splitter, (b) Beam splitter in operation, (c) Beam splitter in a package*
Source: Sche, WOCC, 2004

Full divergence angles as large as 60° are possible for splitting light signals that have a wide range of wavelengths. In most cases, however, the splitter is usually placed at 45° with reference to the source and target platform. If the beam splitter is placed in between an optical path at an angle 45°, a portion of the incident beam energy gets through the splitter and the remaining portion is reflected in a direction 90° from the input beam. Some small portion of the beam, which is passed through the splitter, is absorbed within the beam splitter itself. Because of this feature the beam splitter is considered as an optical window. It has a metallic or dielectric coating on one side with definitive reflecting and transmitting characteristics. The splitters are known as mirror-type beam splitter since some portion of the energy is in fact reflected from the splitter. The typical reflection ratio varies from application to application. Reflection percentages can range from 30% to 80%. A time invariant constant reflection-to-transmission ratio over a large spectral range is desirable. Beam splitters are very sensitive to polarized light due to the reason that the polarization of light is changed when it is incident on a beam splitter oriented at 45°. In a typical mirror-type beam splitter, the S-polarization state is reflected more than the average amount of reflection and the P-polarization state is reflected less than the average when a randomly polarized input light is incidented.

7.6 MICROLENS

Optical lens designs are based on refraction phenomenon. Lens is used to focus the incident light beams. Microlens, a MOEM device, does the same thing. This focusing device is applicable for light beam aligning and optical signal sensing. The microlenses could be a part of a microsystem. An example of microsystem is a micro XY-stage, called micropositioning system (MPS). In many applications, arrays of microlenses are fabricated. Importantly, two types of microlenses are used for various applications. They are

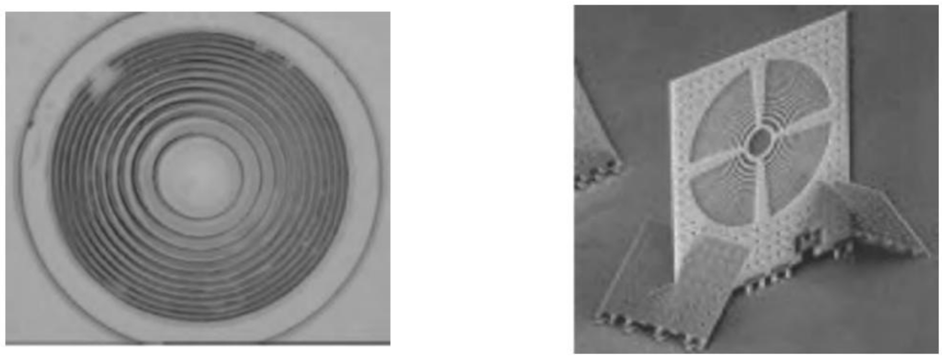
- Micro Fresnel lens
- Micro spherical lens

Typical dimensions of these lenses are in the order of 100–300 μm diameter (Focal length typically 1–3 mm; Fresnel type) and 20–50 μm diameter (Focal length typically 30 μm; monolithic spherical lens), respectively.

7.6.1 Fresnel Lens

Fresnel lens is flat on one side and ridged on the other side. UV-lithography and reactive ion etching (RIE) techniques are used for fabrication. Fresnel lens consists of an array of polysilicon circular rings increasing in width and spacing toward the center (Fig. 7.3 (a)). Since surface micromachining process cannot fabricate curved refracting lenses, the Fresnel lens is the obvious choice. The lens has a plate called lens plate, which is porous and lightweight. The lens plate is usually hinged so it can be flipped up into the light path. A slider integrated with actuator lifts the plate (Fig. 7.3 (b)). The Fresnel lens can also be used to collimate laser light from laser diode.

The micro Fresnel lenses are designed by employing binary optics theory. Binary optics theory is based on diffraction theory. Diffractions theory is more compatible to micro optical components. While conventional lenses, using the principle of refraction, require design parameter in the order of several millimeters of material, the binary optics technology, employing diffractive effects necessitates a fraction of a wavelength in size facilitating the fabrication with micro dimensions and depths as shallow as half a micron or even a nanometer order. Conventional optics lenses are polished to get the desired profile. On the other hand, the binary optics technology uses high-resolution lithography technique to



(a) (b)

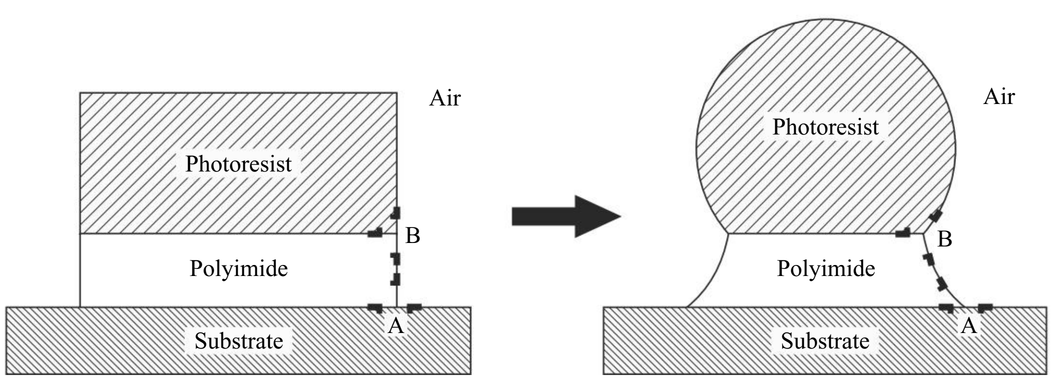
Fig. 7.3 *(a) Photograph of a micro Fresnel lens (diameter 50 micrometer) (b) Out-of-plane free standing micro Fresnel lens with hinge*

transfer a binary surface pattern to the optical element. Binary optics is considered as a new technology capable of manufacturing thinner, lighter lenses and optical devices that were once impossible to make. The pattern is etched into a surface using ion-beam etchers. A single etching step produces a two-level surface, justifying the name binary optics.

Recently developed *gray scale technology* in MOEMS is an advanced technology using which a wide variety of surface profile and shapes with reasonably good optical performance can be achieved. Binary optics rely on a *stairstep* shaped approximation of the ideal surface shape, on the other hand, gray scale can create the smoother shapes, curves, ramps, torroids, or any other shape that is desirable.

7.6.2 Spherical Lens

The micro spherical lens (MSL) is designed by thermal reflow technique. MSL is also called micro-ball lens. Two polymer layers are coated onto a silicon substrate (Fig. 7.4(a)). The upper layer is essentially a photoresist from which the MSL will be created and the lower layer is a polyimide material. The hot air is blown over the two polymer layers. The method of blowing of hot air is known as heat reflow process (HRP). When the upper polymer (Photoresist) is heated beyond its glass transition temperature



(a) (b)

Fig. 7.4 *Micro spherical lenses (*Yang et al., *J of Micromechanics and Microengineering,* 14, 2004*)*

(GTT), the melting polymer will change into a spherical profile in order to minimize the surface energy. The interactive force between the material interfaces caused by surface tension makes it possible the upper profile to form a spherical shape. The polyimide form a pedestal (base) to sustain the ball lens after the HRP is over. The heat reflow process also forms the base polyimide into a trapezoid with arc sides as shown in Fig. 7.4(b).

Because of heat treatment the air tends to minimize the surface energy. The surface energy will be minimized in terms of decreasing the contact area with the substrate. Thus the drop forms a ball shape to maintain small contact area (Fig. 7.5(a)(b)). Let us define the following terms.

E_{ss} = Surface energy of the solid
E_{sl} = Surface energy of the liquid
E_{ssl} = Surface energy of the solid-liquid

The drop will spread (with increasing contact area) onto the substrate when its surface energy reaches the $E_{ss} > E_{ssl} + E_{sl}$ inequality. However, in order to make the melted portion into a ball shape the inequality $E_{ss} > E_{ssl} + E_{sl}$ should be satisfied and as a result of which the melting polymer surface will change into a spherical profile.

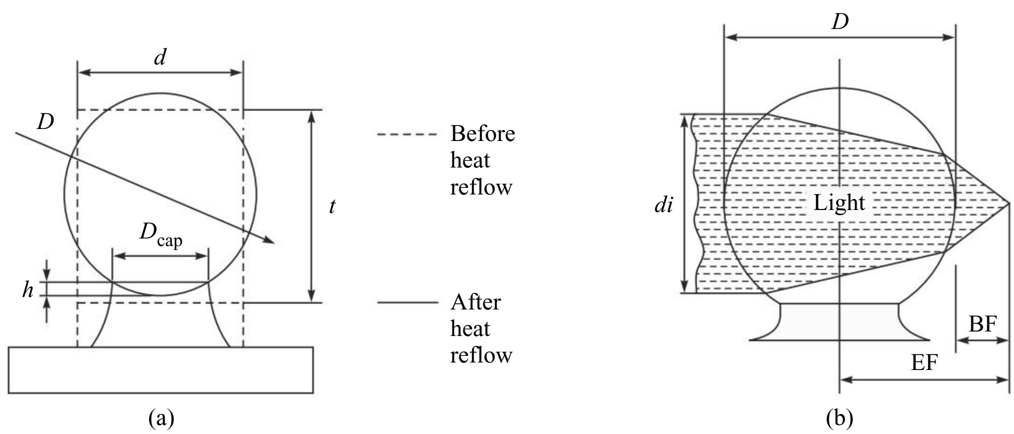


Fig. 7.5 *(a) Volumetric variation; (b) Light refraction in a micro-ball lens (*Yang et al., *J of Micromechanics and Microengineering,* 14, 2004*)*

The mass conservation law can be used to develop the basis for modeling the above phenomenon. The model is primarily used for simulation that is carried out prior to the real design of the microlens. Equation 7.1 and Eq.7.2 give the volumes of the photoresist before and after the heat reflow process, respectively.

$$V_b = \frac{\pi t d^2}{4} \tag{7.1}$$

$$V_a = \pi\left(\frac{D^3}{6} - \frac{Dh^2}{2} + \frac{h^3}{3}\right) \tag{7.2}$$

where d is the diameter of the photoresist, t is its thickness, D is the diameter of a ball lens, h is the height and can be calculated from the following equation.

$$h = \frac{D}{2} - \frac{D}{2}\sqrt{1-k^2} \tag{7.3}$$

$$k = \frac{D_{cap}}{D} \tag{7.4}$$

Here, k is the ratio of the cut cap diameter (D_{cap}) and the ball lens diameter (D).

Ball-shaped MSL arrays are fabricated for many relevant applications. Figure 7.6(a) shows a typical lens arrays. Figure 7.6(b) shows how arrays of lenses are used in an optical connector (called cross-connect optical switch; CCOS), which connects two arrays (input array and output array) of optical fibers, in order to focus the divergent light beams at both the ends. To connect optical fiber array focusing is needed, since the light signal coming from the previous fiber diverse away from its axis of propagation. Similar situation occurs at the receiving end of the fiber array. Therefore, two sets of lens arrays are needed for CCOS.

Example 7.1 It is required to fabricate a ball lens array where the specified cap diameter of each identical ball lens is 2.5 micron. The ration of cut cap diameter to ball lens diameter is 0.5. Calculate, (i) the height up to which the lens cannot be used, (ii) the volume of each lens after the heat treatment. For a 16 × 16 configuration how much polymer would be needed?

Solution: The cut cap diameter of identical lens to be put into place is 2.5 microns.

The ratio of cut cap diameter to ball lens diameter, $k = 0.3$

So, the ball lens diameter, $D = 2.5/0.3 = 8.33$ microns

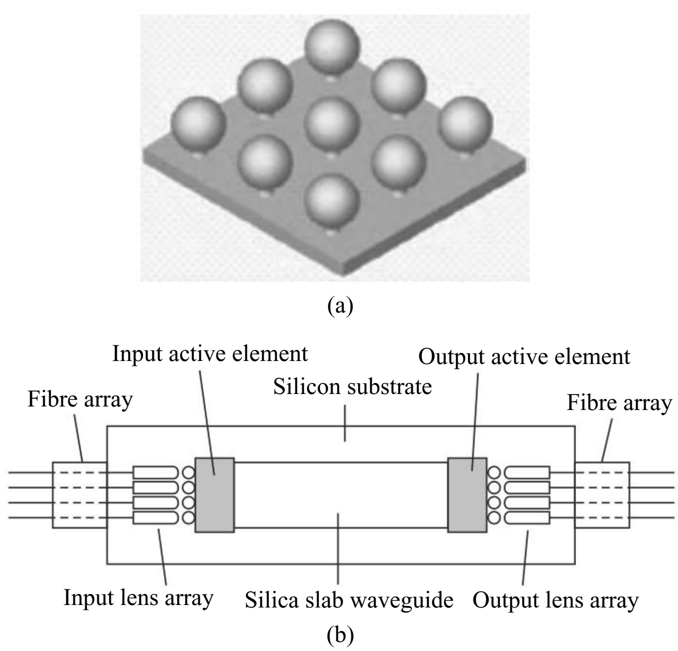


Fig. 7.6 *(a) Microspherical lens arrays, (b) An example of application of microlens array in cross-connect optical switch*
Source: Alexei et al., *JM*^{3} *2(4)*, 2003, SPOIE

The height up to which the lens cannot be used is,

$$h = \frac{D}{2} - \frac{D}{2}\sqrt{1-k^2} = \frac{8.33}{2} - \frac{8.33}{2}\sqrt{1-0.09} = 0.192 \text{ micron}$$

The volume of each ball lens after the heat treatment can be found out from the above equation,

$$V_a = \pi\left(\frac{D^3}{6} - \frac{Dh^2}{2} + \frac{h^3}{3}\right)$$

$$= \pi\left(\frac{8.33^3}{6} - \frac{8.33 \times 0.192^2}{2} + \frac{0.192^3}{3}\right) = 302.1653\ (m\mu)^3$$

For a 16 × 16 configuration the total number of lenses will be 256. So the total volume of polymer = 256 × 302.1653 = 77354.31 $(\text{m}\mu)^3$.

7.7 MICROMIRRORS

Micromirrors are found in many precision equipments and communication systems. Micromirrors are circular, rectangular or square in structures. When light falls upon the micromirror it is reflected. It can be used as reflector, a switch in optical signal transmission system, a feedback sensory element in position control system, and so on.

By utilizing electrostatic actuation method the micromirror is rotated or tilted to a pre-defined state to achieve controlled optical signal transmission (Fig. 7.7). The angle of rotation could vary from 1^o to 45^o depending upon the application requirement. In position control system the reflected light is observed and calibrated in terms of displacement or angle, which could be proportional to the input physical parameter, for instance, vibration magnitude of a rotary shaft. Optical switching through micromirror is achieved by integrated actuation mechanism.

Micromirror is usually hinged in the base and can be tilted by applying electrostatic force. Any amount of tilting is achieved through the application of equivalent electric potential (Fig. 7.7). Figure 7.7 shows a SEM (Scanning Electron micrograph) of 2 degree-of-freedom (2-DOF) micromirror. The mirror is called torsional micromirror. Figure 7.7(c) is the close-up of torsion hinge suspension and force redirecting linkage. This linkage provides roughly 5 times the range of motion of the unmodified thermal actuators. The mirror consists of two large plates approximately 500 μm × 500 μm × 4.25 μm. The right-hand panel is a support structure, which rotates on substrate hinges on an axis in the plane of the substrate. The left-hand panel is the actual mirror, and is suspended from the support structure by two U-shaped torsion springs. The mirror is constructed by sandwiching a 0.75 μm oxide layer between the 2 μm thick poly-1 layer and the 1.5 μm thick poly-2 layer.

Micromirrors are produced as single element containing a single micromirror or a group of micromirrors in a single wafer in a two dimensional array. The array type micromirrors are primarily used for scanning and image display applications.

7.8 DIGITAL MICROMIRROR DEVICE (DMD)

Digital Light Processing (DLP) is a technology and DLP™ is the trademark of Texas Instrument. The device capable of implementing the DLP technology is called Digital Micromirror Device (DMD™).

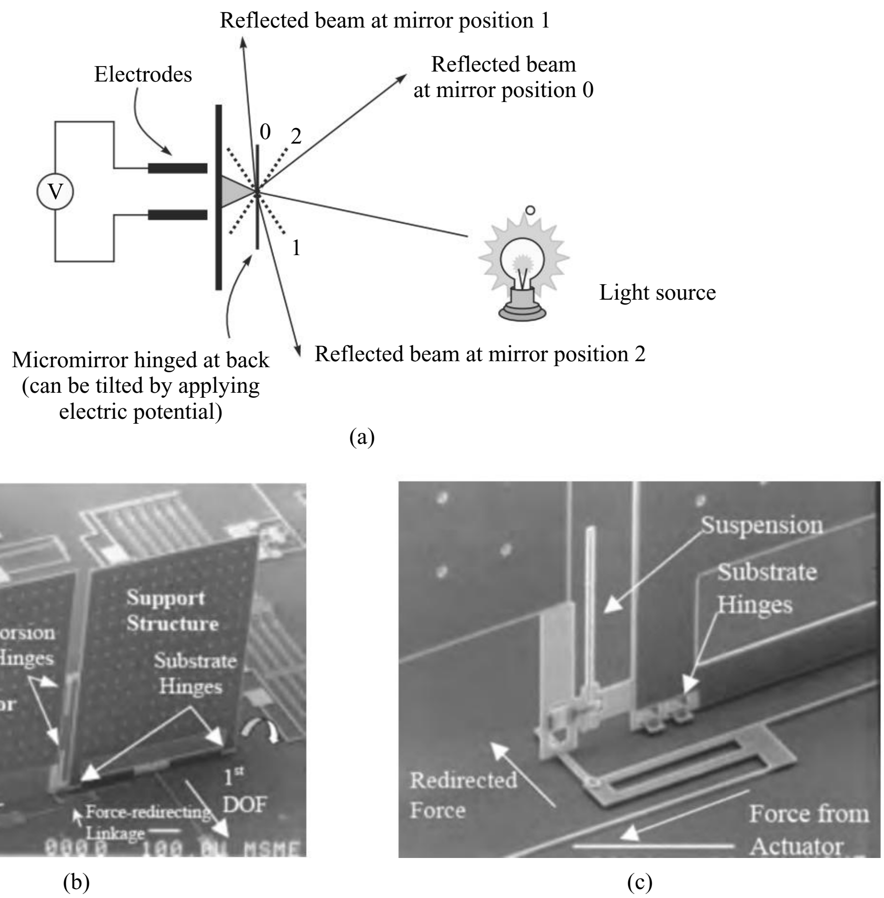


Fig. 7.7 *(a) Principle of operation of a typical micromirror, (b) SEM (Scanning Electron micrograph) of the back side of 2 degree-of-freedom mirror, (c) Close-up of torsion hinge suspension and force redirecting linkage. This linkage provides roughly 5 times the range of motion of the unmodified thermal actuators*

Source: Last and Pister, *MOEMS '99,* Germany

DMD™ is a micromechanical silicon chip. From operational point of view DMD is a precision light-switching device, capable of modulating the light digitally by the use of huge numbers of microscopic mirrors, which are arranged in a definitive manner, usually in the form of rectangular array. The separation between the micromirrors in the array is in the order of micrometer. The device is in fact a Spatial Light Modulator (SLM), primarily used for imaging applications.

The micromirrors in the device are reflective elements and their shapes are flat, rectangular and polished. Approximately more than one quarter of a million numbers of mirrors (typically 640*480 pixels) is fabricated into a single device. The device is a unique combination of optical, mechanical and electrical systems and is much more than a typical IC (Integrated Circuit) chip. The functions of the micromirrors are to:

- Direct the light signals toward a dedicated pixel space
- Switch the light signal in the order of thousand times per second

7.8.1 Principle of Operation of the Device

The light source (lamp) is modulated in the DMD chip. Note that the lamplight is a constant light source. Its intensity has to be modulated in response to incoming video signal. After being modulated the light falls on the dedicated pixel space. Each mirror corresponds to a single pixel in the projected or displayed image (Fig. 7.8(a)). The 3D construction is such that the mirrors are capable of switching into ON and OFF state. When it is ON, the mirror reflects light incoming from the lamp source. External circuit controls the mirrors. Each mirror is hinged and interfaced with a CMOS RAM memory block, accountable to store/provide display-specific information. The size of RAM depends on the number of mirrors present in the DMD. The entire block of RAM is divided and allocated to each mirror in order to store three most important display-specific parameters (DSP) such as address of the mirror, the tilting direction (ON or OFF) and the timing of tilting. The interfacing circuits, RAM block and mirror arrays are extremely complex. For the purpose of understanding Fig. 7.8(a–b) illustrates a schematic view of such a device, although a DMD™ chip is entirely different from this representation.

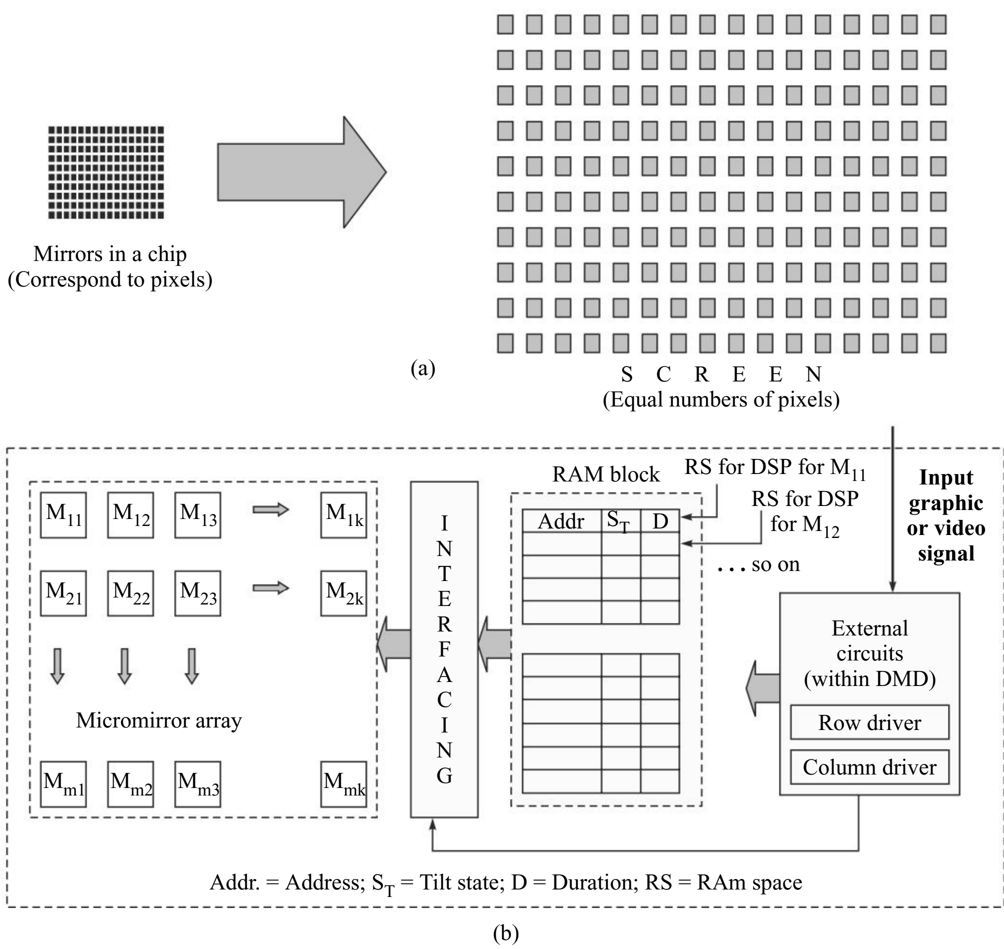


Fig. 7.8 *(a) Mapping between the pixel space on the screen and the DMD mirrors, (b) Schematic diagram of an array of micromirrors for display applications*

There are row drivers, column drivers and timing circuitry built into the device. Applying the electrostatic force through microelectrodes constructed into the mirror can tilt an addressable mirror (For reference see Fig. 7.7(a)). A particular mirror is made either ON or OFF by addressing it through RAM memory in terms of applying electrostatic voltage to the microelectrodes. The video signal to be displayed is available as equivalent electrical signal. The incoming video signal has to be decoded in such a way that the electrodes, which control each micromirror, must be activated at the appropriate time with accurate duration of tilt. The incident light from the lamp is reflected from each micromirror and then amplified by the projection lens and directed to the pixel space (screen). Tilting direction and the duration of tilt assures modulation of incident lamplight.

7.8.2 How it Works

Figure 7.9 illustrates a schematic diagram of a complete image display system that uses the DMD. The display system consists of the following nine units.

- Projector lamp
- Condensing lens
- Color filter (wheel)
- Shaping lens
- Processor
- Memory
- DMD
- Projection lens
- Screen

The projector lamp is a source of light that is focused onto the DMD. The light from the lamp (source) is passed through the condensing lens. The condensing lens focuses the light toward the DMD. The color filter is a transparent wheel consisting of three basic colors. The white light from the lamp is modulated into colored light by this filter and made to fall on the micromirrors of the DMD device through the shaping lens. The wheel spins illuminating the DMD sequentially with red, green, and blue light. The video signal to be displayed is fed to the DMD. The mirrors are turned ON depending on

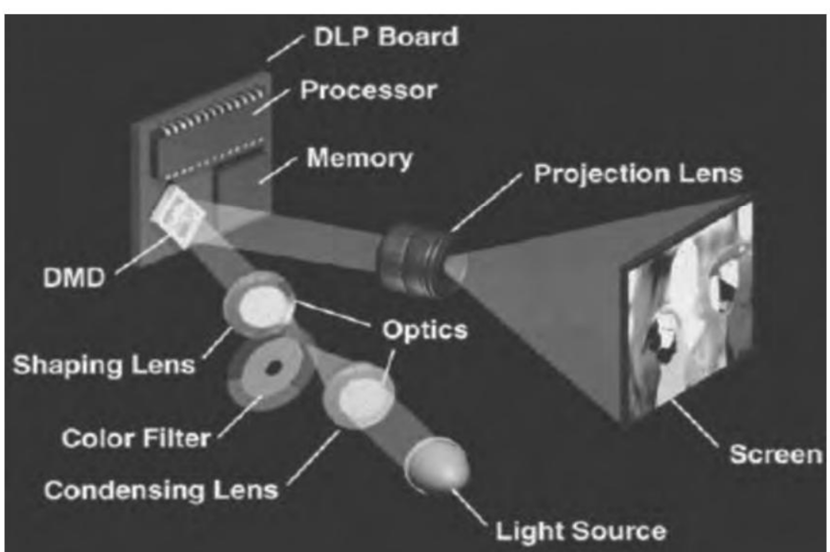


Fig. 7.9 *The basic units of a DMD™-based display system*
Courtesy: Texas Instruments

how much of each color is needed for displaying. DLP board consists of three units, namely the processor, memory and the arrays of micromirrors. The electrical video signal is decoded and fed to the corresponding electrode pairs that control the micromirror for tilting. Further, the duration of the ON-OFF timing of each micromirror determines the level of intensity (called grayscale) of the pixel. The level of intensity in the pixel is in turn related to the color of the video signal. Grayscale output is achieved by pulse width modulation (PWM). Brighter pixels correspond to longer pulses, darker pixels to shorter pulses. The reflected light from each micromirror is then amplified by the projection lens and splashed onto the front or back of the screen. The display of image on the back of the screen corresponds to rear-projection set.

A '1' or '0' in the RAM cell tilts the mirror ±10 degree. That is a '1' in the memory causes the mirror to rotate +10 degrees, while a '0' in the memory causes the mirror to rotate –10 degrees. A mirror rotated to +10 degrees reflects the lamp light into the projection lens and the pixel appears bright at the screen, where as the mirror rotated to –10 degrees misses the projection lens and the pixel appears dark. Rapid switching of the mirrors is a desirable feature. The settling time is very fast (approximately 16 micro second) and far more than the normal human eye. The input video data rates and data bus widths are designed and specified in such as way that the memory-mirror array must be refreshed 48–60 times during a single video frame so as to enable to provide a display with 16 million possible colors. One can note that the system can produce more than 16 million colors. The switching speed is very much related to the number of colors it can display[1].

The size of DMD chip is approximately 0.625 inch. The size and fill factor of a single mirror is about 16 μm^2 and 90%, respectively. Fill factor is the ratio of reflective area to the actual area of the micrometer. The three-chip based projection systems are called high-definition system. This system consists of three high-resolution DMDs. Each DMD is for the one primary color of red, green or blue (RGB).

Micromirror platform can be used for many other applications including beam positioning, optical cross connect switches, VOA (variable optical attenuators), OADM (optical add/drop multiplexer), tunable laser, portable display and HDTV applications.

7.9 LIGHT DETECTORS

Diffraction is a phenomenon, which implies that the wavefront of a propagating light wave bends in the neighborhood of obstacles.

Let us discuss the diffraction phenomenon based on single-slit (Fig. 7.10(a)) and double-slit (Fig. 7.10(b)) interference. If we consider diffraction through a single slit then the properties of the system are completely dependent on the ratio λ/a, where λ is the wavelength of the light and a is the aperture of the slit. The mapping of the intensity pattern some distance away from the slit consists of bright and dark fringes. The angle at which the dark fringes (minima) occur is given by,

$$\sin\theta = \frac{m\lambda}{a} \qquad m = 1, 2, 3\ldots$$

$$\text{But,} \qquad \theta \approx \theta' \approx \sin\theta \approx \tan\theta = \frac{y}{D} \tag{7.5a}$$

[1] To understand the principle of display technology the readers can read the introduction chapter of the book Monochrome and Color Television by Gulati.

So, $$y \approx \frac{m\lambda D}{a}$$

where, y is the displacement or deflection at which minima occur, D is the distance at which observation is carried out, λ is the wavelength of light (refer Fig. 7.10(a)). Note that the intensity pattern is

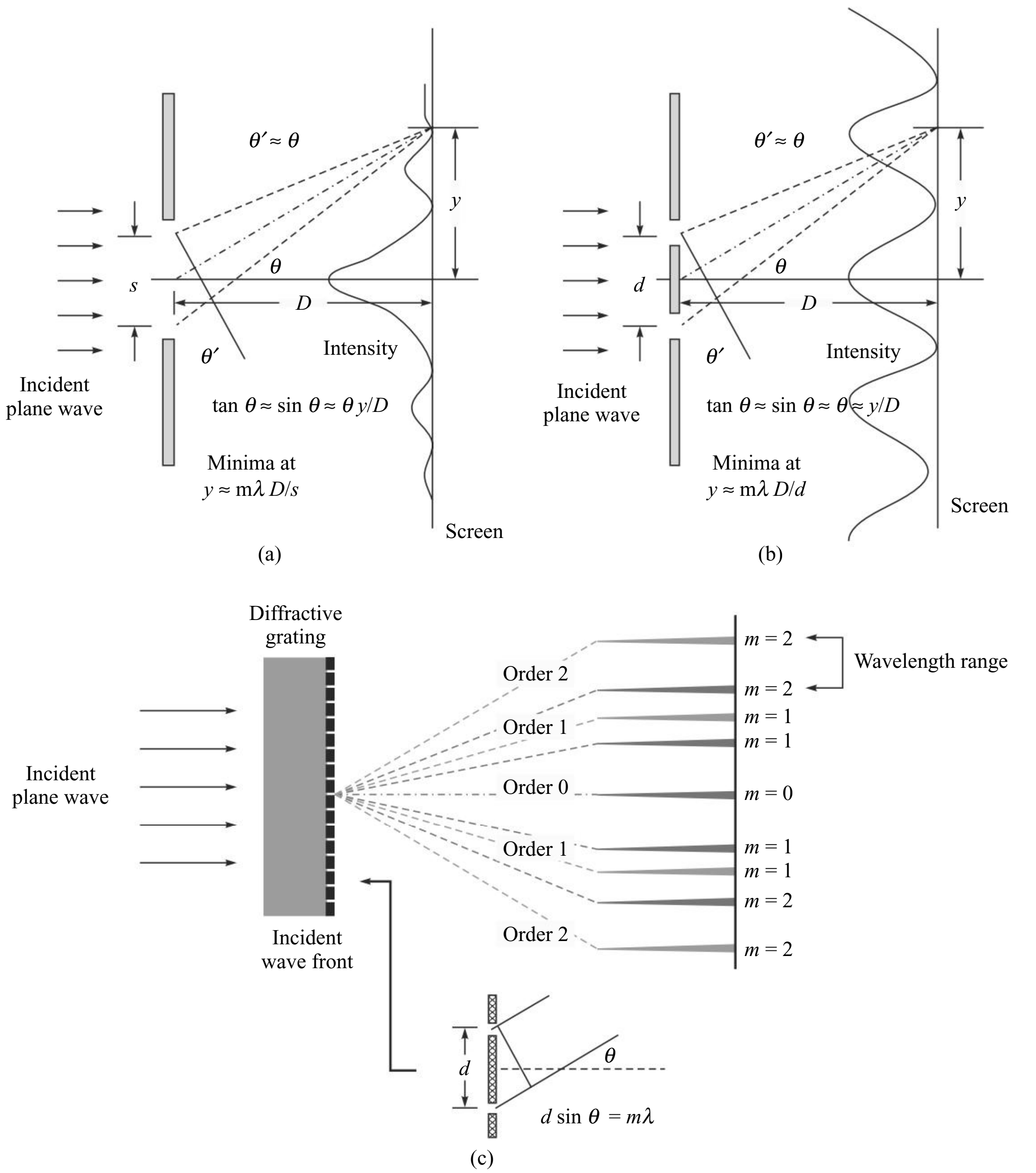


Fig. 7.10 *(a) Single-slit example, (b) Double-slit, (c) Diffraction grating showing the wavelength separation*

determined only by the ratio λ/a. In the middle ($m = 0$), a central bright fringe will be observed and it is the largest bright fringe.

For double slit configuration, the situation is different in the sense that the condition for maximal points (instead of minima as in case of single slit) is given by,

$$y \approx \frac{mD\lambda}{d} \tag{7.5b}$$

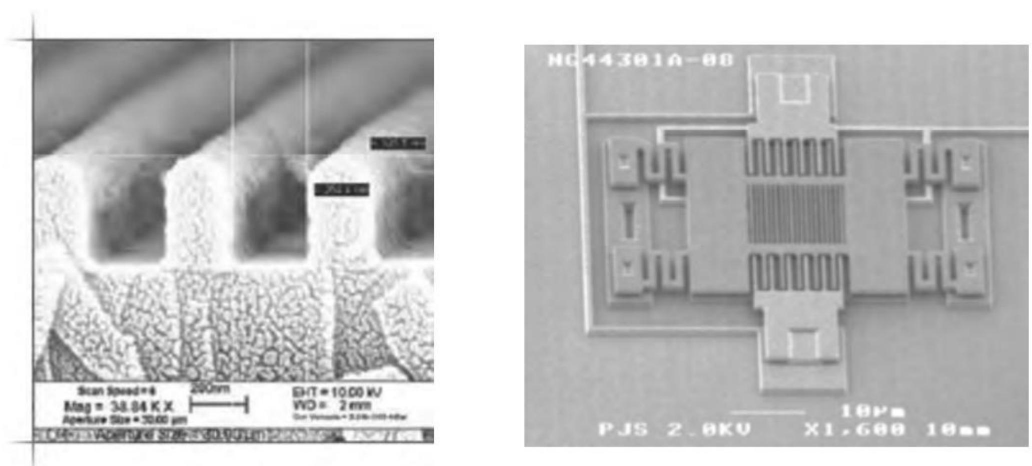
where, d is the separation between the slits or grooves (Refer Fig. 7.10(b)).

A diffraction grating (DG) is a set of parallel slits (equivalent to obstacles that are placed alternatively) used to disperse or bend the light (Fig.7.10(c)). In DG there exist many slits. Since, diffraction grating deals with the light signal it is an optical device. It is primarily used to observe the different wavelengths or colors contained in a beam of light. They are also used to separate the dense wavebands. The gratings allow efficient diffraction at *grazing* incident angles.

There are specific directions along which the light waves from the different openings interfere with each other in phase and in these directions the light intensity is observed as maximum. The specific directions along which the intensity will be maximum depend on the wavelength of the light waves. Since the light is composed of many waves having different wavelengths, the direction of propagation will be different for each wavelength. As a result a specific wavelength can be separated or detected by the use of diffractive grating, hence the name light detector. The intensity of light at the screen is proportional to the number of slit in the grating.

MEMS technology based diffraction gratings are rather called nano diffraction grating (NDG). NDGs are obviously MOEM devices. The device consists of thousand of narrow and closely spaced parallel slits (Fig. 7.11). The NDG shown in the figure has a depth and width of 395 nm and 294 nm, respectively.

Example 7.2 Consider a single slit flap. For a slit of width 3 micrometer and light wavelength 500 nanometer at order $m = 2$ on a screen distance of 1 cm, calculate the displacement from the centerline for minimum intensity.



(a) (b)

Fig. 7.11 *(a) Dementias of a typical diffraction grating – High scale view, (b) Global view*
Courtesy: MEMS Optical Inc.

Solution: Given: Slit width, $d = 3\ \mu m = 3 \times 10^{-6}$ m
Wavelength of light, $\lambda = 500$ nm $= 500 \times 10^{-9}$ m
Order, $m = 2$
Screen distance, $D = 1$ cm $= 10^{-2}$ m
Putting the values in the corresponding equation we have,

$$y \approx \frac{mD\lambda}{d} = \frac{2 \times 10^{-2} \times 500 \times 10^{-9}}{3 \times 10^{-6}} = 3.33\ \textbf{mm}$$

Example 7.3 Consider a double single slit configuration. Determine the wavelength of light when the following parameters are known.

Slit separation distance (center to center), $d = 4.9\ \mu m = 4.9 \times 10^{-6}$ m
Order, $m = 2$
Screen distance, $D = 2$ cm $= 2 \times 10^{-2}$ m
Observed at a point 0.55 cm from the center line, i.e. $y = 0.46 \times 10^{-2}$.

Solution: The approximate displacement formula is written as,

$$y \approx \frac{mD\lambda}{d}$$

$$\Rightarrow \qquad \lambda = \frac{yd}{mD}$$

Putting the value given above,

$$\lambda = \frac{0.46 \times 10^{-2} \times 4.9 \times 10^{-6}}{2 \times 2 \times 10^{-2}} = 563.5 \times 10^{-9} \text{ m} = 563.5 \text{ nm}$$

7.10 GRATING LIGHT VALVE (GLV)[2]

Another spatial light modulator technique known as Grating Light Valve™ (GLV™) has taken momentum in image display applications. Unlike DMD™ technology, GLV™ technology works on the principle of diffraction. The GLV™ device consists of parallel rows of tensile reflective membrane type micro-ribbons as shown in Fig. 7.12. The ribbons are rigidly mounted on the base at the longitudinal ends and are placed equidistantly. The ribbons are suspended over a thin air gap, which exists between the ribbons and the wafer. The length of the ribbon is approximately 20 μm and the air gap is approximately 1300 Angstroms.

For monochrome display the width of the ribbons called pitch, are same but for color display the pitch varies depending upon the wavelength. The pitch corresponding to red color is higher than that of green, which is higher than that of blue. The ribbon pitch is approximately 5 μm. Other dimensions of GLV™ ribbons are typically 100 μm long or less and 100 nm thick.

Silicon nitride, ceramic material is used to fabricate the ribbons. The ceramic material can provide reasonably good tensile strength and durability. An aluminium coating is formed on the top of the ribbon in order to get smoothed surface. The rough surface may scatter the light reducing the efficiency. The

[2] Silicon Light Machines, Grating Light Valve (GLV) and GLV are trademarks of Silicon Light Machines.

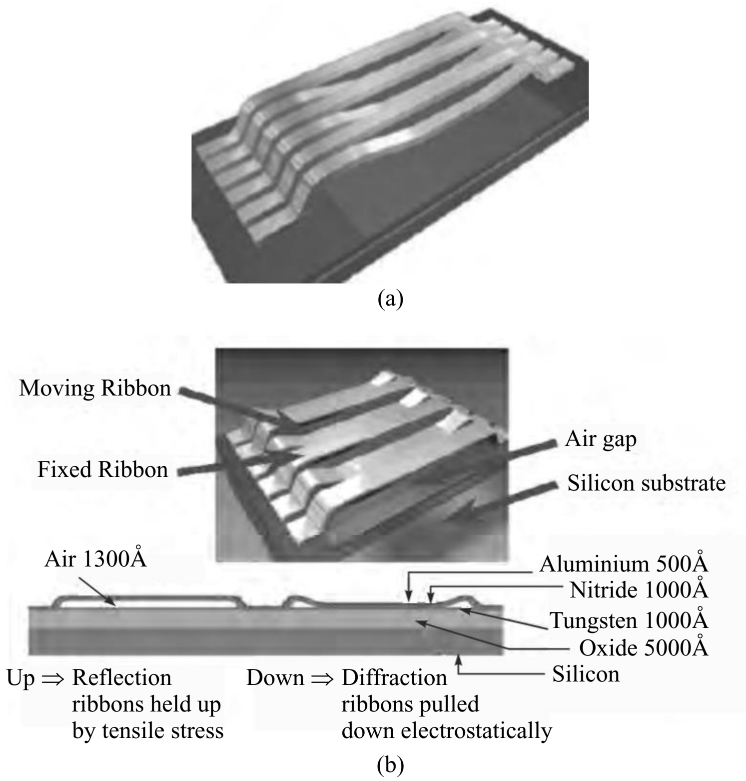


Fig. 7.12 *(a) A typical GLV device with reflective micro-ribbons; (b) Dimensions of various layers of a typical GLV*
Courtesy: Silicon Light Machines, Sunnyvale

aluminium layer is extremely thin. The coating also behaves as a conductor but not necessarily reflective (Why?). Microfabrication technique sequences such as photolithographic masking, deposition, etching, metallization, etc. are employed to construct the device. We like to describe the principle of operation of monochrome display at the beginning and then will be discussing on how colored images are produced by the use of GLV™ technology.

7.10.1 Diffracting Mode of Operation

Normally all the ribbons are flat and straight. When all the ribbons are flat the device is said to be in the reflective mode. Applying electrical voltages can pull the ribbons vertically down relative to the underlying surface. In practice, alternative ribbons are pulled down in order to make the device a diffraction grating. This is called diffractive mode of operation of the ribbon. The amount of deflection, i.e. pull-down distance of the ribbon, should be comparable to the wavelength of light so that the phenomenon of diffraction can be best utilized. Typically the amount of deflection is zero to one quarter of the wavelength of the green light (i.e. $\lambda/4$, where λ is the wavelength of the green light). The applied voltage controls the amount of downward deflection of the ribbons.

One should note that, when the device is in reflection mode (flat state) the light gets reflected and goes back in the same direction it had come. This implies dark image that is produced in the viewing

system (screen). If a particular ribbon is deflected down, the incident lights will be diffracted due to the reason that the reflected light from the deflected ribbon will not be in phase with the un-deflected ribbon adjacent to it. Further, since the deflection and the pitch of the ribbons are in the order of wavelength of the incident light, the reflected light from the deflected ribbon will see a slit in front and will bend that depending on the amount of ribbon deflection and the width of the ribbon. Since each ribbon can be pulled down a controlled distance into the air gap by means of an electrostatic voltage, by pulling down the alternative ribbons from their flat state to the deflection state, a square-well diffraction grating can be formed. Light waves reflecting off adjacent up and down ribbons will be out of phase with each other. Since the light is composed of many waves having different wavelengths, the direction of propagation will be different for each wavelength. In other words, the deflection effect will cause the waves to interact in a way that each wavelength of light will be radiated at a different angle. The desired angle at which certain wavelength and therefore colors of light are to be collected is achieved by varying the pitch (hence the slit aperture) and the degree of pull-down of the ribbons. It is as good as saying that a desired color can be obtained by varying the pitch and deflection parameter. This diffracted light can produce a bright spot in a viewing system.

7.10.2 Building Blocks

The above principle is employed in order to display the image. Further description follows. The input video signal is first formatted in terms of converting it into a coded digital signal. The signal is then fed to a digital driver, which is interfaced directly with the GLV™ device (Fig. 7.13). Note that for black and white display the pitch of all the ribbons are same. But for color display three types of ribbons with different width are needed. The respective width corresponds to the three primary colors such as Red, Green and Blue (RGB). As already mentioned the width of the ribbon is proportional to the wavelength of the light. Since the wavelength of the red light is greater than that of green light the pitch width of the ribbon corresponding to the red light should be greater than that of the green light so that the green, red and also the blue light can be directed to travel along a common direction, i.e. toward the viewing system or screen. The white light falls on the GLV™; it then gets diffracted dividing into three components (RGB) and travels in parallel towards the display screen (Fig. 7.13(b)). Had the pitch width not been different, it would have been impossible to direct these primary light wavelengths (RGB) into same direction.

Figure 7.13(c) illustrates a more practical approach as far as realization of GLV™ technology is concerned. The arrangement can be considered as a light machine. There are seven components in this arrangement.

- White light source
- Lens (two numbers)
- RGB color wheel
- Digital driver
- Turning mirror
- Screen
- GVL device

The light source produces white light. The two lenses are placed at appropriate places in order to focus the light in proper direction: one is placed in front of the GLV™ device and another is placed back to the turning mirror. RGB wheel filters the white light to produce the three primary components of color (wavelength). The filtered light red, green or blue is allowed to fall on the reflecting turning mirror. The turning mirror performs two functions. Firstly, it directs the light onto the GLV™ device and secondly it works as an optical stop that blocks the reflected light from the GLV™. The reflected light from the turning mirror is now directed toward the GLV™ device, which is intended for diffraction. The

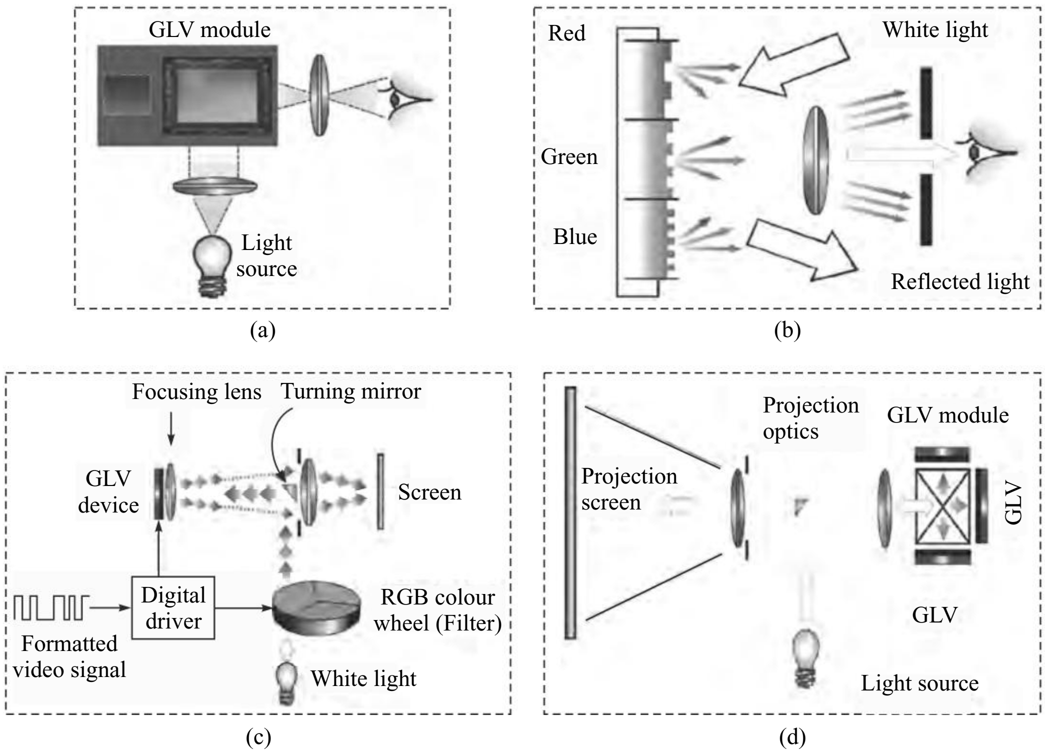


Fig. 7.13 *(a) A simple GLV based projection; (b) The red-green-blue triad; (c–d) A more practical approach of image display*
Courtesy: Silicon Light Machines, Sunnyvale

filtered and diffracted primary colored light then travels horizontally towards the screen for illusion after being focused through a focusing lens.

7.10.3 Detail Description

The ribbons are placed vertically in the GLV™ device. Figure 7.14 shows a hypothetical grating valve, which has 4000 ribbons in one line scan. The entire valve has been segmented into 2000 lines. The total area has been divided into 2000000 numbers of pixels. The width of the ribbon could be 5 micrometer. A pixel consists of four ribbons placed vertically. The ribbons in a pixel are deflected (controlled), by applying voltage to them. The image element in the screen corresponds to a pixel in the valve device. The formatted video signal coming from digital driver is fed to the GLV™ as well as to the color wheel for synchronization. This is an essential requirement for image projection. By synchronizing the image data stream's primary red, green and blue pixel data with the appropriate filtered white light, combinations of red, green and blue diffracted light is directed to the screen. The GLV™ device can work on a vertical pixel line, which are scanned left and right to produce millions of pixel image with extraordinary resolution.

The principle of operation is based on the creation of a dynamic, tunable grating that precisely varies the amount of light that is diffracted or reflected (Reflection implies dark image). The formation of grating is dynamic due to the reason that the deflection of the ribbons in a pixel depends on

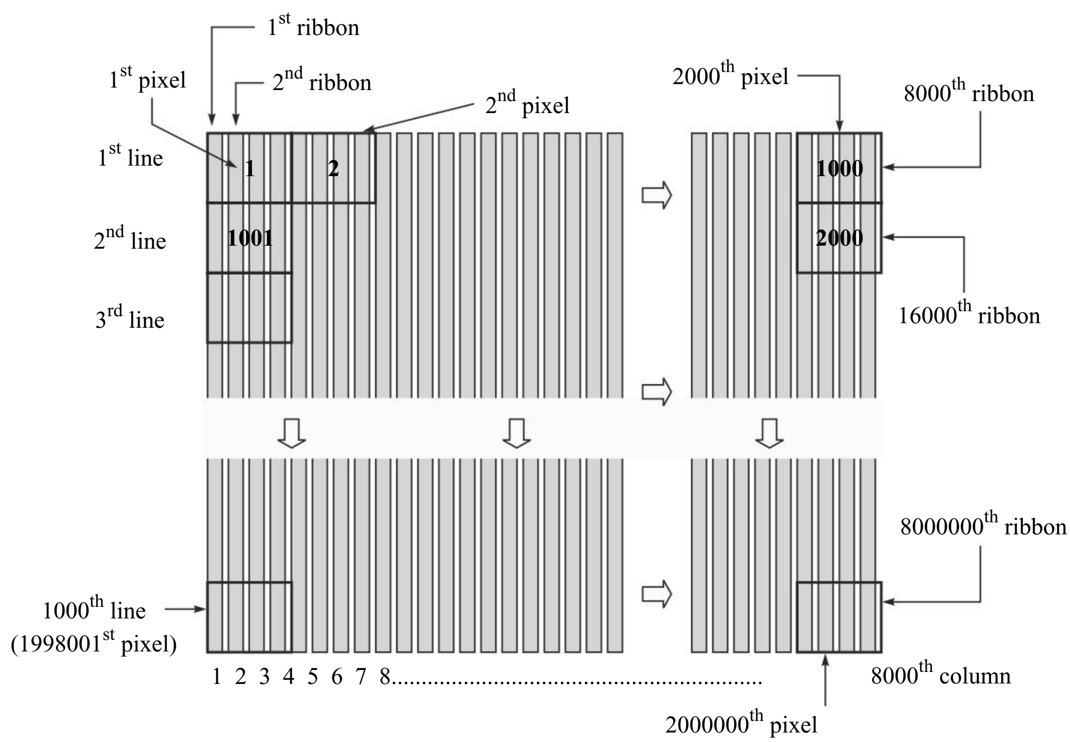


Fig. 7.14 *A schematic diagram of a hypothetical grating valve*

corresponding video signal, which is considered to be time variant. The GLV™ device is essentially an addressable dynamic diffraction grating since the ribbons in a pixel are controlled in terms of addressing them by the use of electronic circuitry.

The pixels are called controllable picture elements. The white light is selectively diffracted to produce an image of bright and dark pixels. As can be seen from the figure, each pixel is built with multiple ribbons. Multiple ribbons will produce greater image brightness. By activating (deflecting) more numbers of ribbons in a pixel, superposition of diffracted will occur resulting *constructive crest.*

7.10.4 A Faster Device

If the up and down deflection switching of the ribbon is made faster then modulation of the diffraction will produce many gradations of grayscale (colors). Essentially, the ribbon switching occurs very quickly in the order of 20 nanoseconds. Therefore, this technology is very fast settling and response time as compared to the LCD-based display technology. Even the GLV™ technology is faster than the DLP™ technology. The high up and down switching is essential in order to achieve a wide range of gray or color variations. More faster switching can be achieved by reducing the mass of the ribbons. One can note that GLV™ switching speeds can make it easy to implement an 8-bit or greater gray scale, and are fast enough to support colors and grays.

7.10.5 Summary

A more effective and efficient color projection system can be built using three GLV™ devices, corresponding to three primary wavelengths (colors such as R, G and B). In this case the white light is

passed through color separators or filters. The filters are called dichroic filters and are mainly used for a variety of applications where it is necessary to isolate certain regions of the visible spectrum such as red, blue and green light. Then the filtered light is incident on three separate GLV devices, respectively. The diffracted light is collected from the respective GLV device and directed through the optical system (lens) to a viewing screen. In summary it can be said that SLM makes tiny silicon machines for handling light. SLM uses Grating Light Valve (GLV) technology to build MOEMS onto ordinary silicon semiconductor wafers.

7.10.6 Modeling

The spring and the capacitor can describe the electromechanical model of the GLV. The vertical downward Coulomb force exerted on the addressed ribbon because of the applied voltage based on formatted video signal is given by,

$$F_{\text{elec}} = \frac{\beta V^2}{(h-\delta)^2} \tag{7.6}$$

where, F_{elec} is the Coulomb force, β is a parameter that depends on the ribbon area and the effective permittivity, V is the applied voltage, h is the gap height and δ is the deflection. The equilibrium will occur when the Coulomb force equals the Hooke's Law force given by,

$$F_{\text{mech}} = k\delta \tag{7.7}$$

where, F_{mech} is the Hooke's force, k is called Hooke's constant. Now, at equilibrium $F_{\text{elec}} = F_{\text{mech}}$. From the above equations it can be written that,

$$\frac{\beta V^2}{(h-\delta)^2} = k\delta$$

$$\Rightarrow \qquad \delta^3 - 2h\delta^2 + h^2\delta - \frac{\beta V^2}{k} = 0 \tag{7.8}$$

At low voltage, the above equation does not have a stable solution for $\delta > h/3$. This implies that the Hooke's force is not strong enough to balance the Coulomb force and the ribbon will snap to the base (Refer Fig. 7.15).

A simplified practical formula for the deflection, which is a function of the applied voltage is however, given by,

$$\delta(V) = \frac{h}{3}\left[1-\left\{1-\left(\frac{V}{V_s}\right)^{w}\right\}^{2/3w}\right] \tag{7.9}$$

where $w = 1.8$ called fitting parameter and V_s is the saturation voltage given by, $V_s = (4kh^3/27\beta)^{1/2}$.

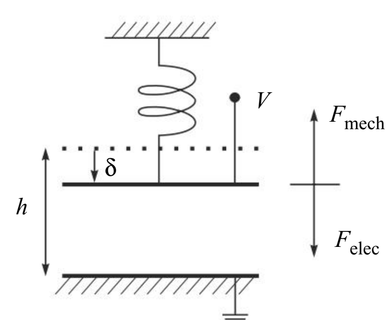


Fig. 7.15 *The spring and the capacitor model of the GLV ribbon*

The diffraction phenomenon of the GVL can be described mathematically. Mostly, scalar diffraction theory is used as a good description for the GLV diffraction. The diffraction is usually characterized by the efficiency factor. Here the diffraction efficiencies for 0^{th} and 1^{st} order light are given. They are,

$$\eta_0 = R_r\mu^2 \cos^2\left(\frac{2\pi\delta}{\lambda}\right) + R_g(1-\mu)^2 + 2\sqrt{R_r R_g}\,\mu(1-\mu)\cos\left(\frac{2\pi\delta}{\lambda}\right)\cos\left(\frac{2\pi(\delta-h)}{\lambda}\right)$$

$$\eta_{\pm 1} = \frac{4}{\pi^2}R_r \sin^2\left(\frac{\pi\mu}{2}\right)\sin^2\left(\frac{2\pi\delta}{\lambda}\right) \tag{7.10(a)(b)}$$

where, R_r is the ribbon reflectivity, R_g the gap reflectivity, μ is called fill-factor of the ribbon. Maximum diffraction is achieved when the ribbon deflection is quarter wavelength.

Example 7.4 From the simplified formula presented in Eq. 7.9 plot the relationship between the applied voltage, V versus deflection, δ in electromechanical model of the GLV. The following assumption and parameters are given.

(i) $\delta < h/3$.
(ii) $V_s = (4kh^3/27\beta)^{1/2}$
(iii) Fitting parameter, $w = 1.8$
(iv) Hookes constant, $k = 4000$ N/m
(v) β that depends on the ribbon area and the ε_0, $= 23.7 \times 10^{-12}$ units (mks)
(vi) Gap height, $h = 100$ micron.

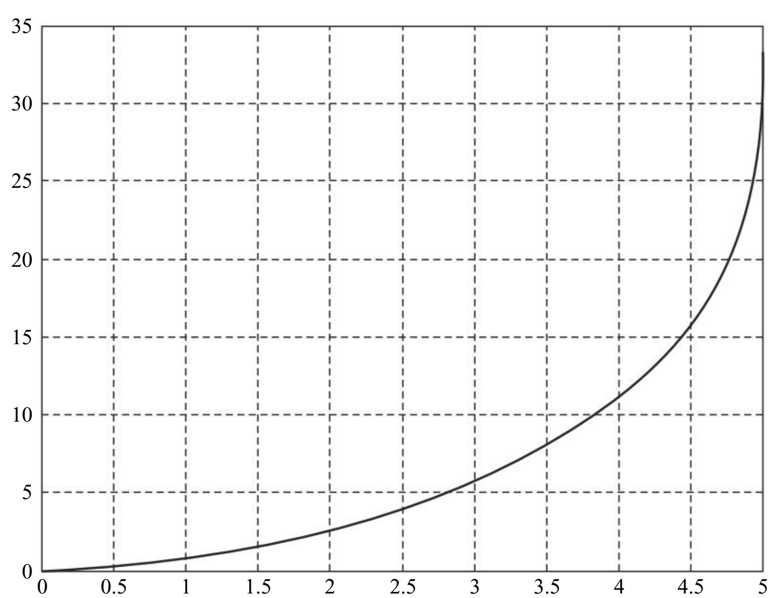
Solution: Now, the saturation voltage, $V_s = (4kh^3/27\beta)^{1/2}$

$$= [\{4 \times 4000 \times (100 \times 10^{-6})^3\}/(27 \times 23.7 \times 10^{-12})]^{1/2}$$

$$= 5 \text{ volts}$$

Now, putting the values of V_s, w and h in Eq. 7.9, we have,

$$\delta(V) = \frac{100\times 10^{-6}}{3}\left[1-\left\{1-\left(\frac{V}{5}\right)^{1.8}\right\}^{2/3\times 1.8}\right] \approx 33.33\left[1-\left\{1-\left(\frac{V}{5}\right)^{1.8}\right\}^{0.37}\right] \text{ (in micrometer)}$$



7.11 OPTICAL SWITCH

Mirror-based MOEMS optical switches have drawn attention in optical network since long. The switch is designed for regulating light beams. By using the switch light can be directed in the same direction or in another according to the requirement. There are mainly two types of switches based on the way the mirrors are activated. They are interception type and rotation type. The former type uses multiple mirrors requiring large linear actuation and the design is very complex. The latter type of switch is called tortional micromirror type and uses single mirror requiring large angular actuation (Fig. 7.16).

The micromirrors can route light signal from and to any input/output optical fibers of 2D or 3D array (Fig. 7.16(c)). This suggests another way of classifying the switches in terms of the number of input-output ports present. Accordingly, there are four types of switches, namely,

- Single input single output
- Single input multiple output
- Multiple input single output
- Multiple input multiple output

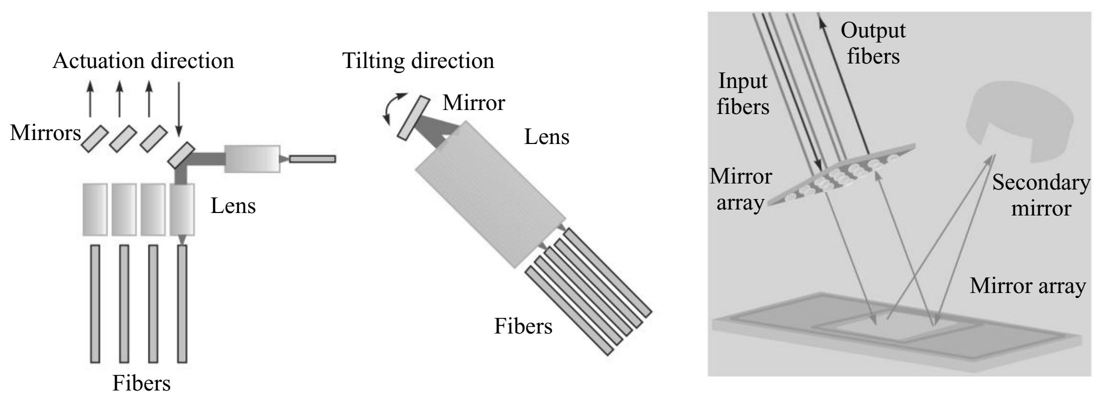


Fig. 7.16 *Optical switch; (a) Beam interception type; (b) Beam rotation type; (c) Light can be directed in any direction*

Regardless of their types the switches are activated using the principle of electrostatic actuation. Magnetic actuation based optical switch will be discussed in the next chapter. Figure 7.17 shows a simple 2D single input multiple output optical switches. It has one input port and four output ports. The incoming light signal can be directed to any of the output port that is connected to the optical fiber through the microlens arrangement. If the light signal is required to be directed to the optical fiber number 2 then the corresponding motor M_2 must be activated by the electrostatic actuation (Fig. 7.17 (b)). Similarly, if the light signal is required to be directed to the optical fiber number 4 then the corresponding motor M_4 must be activated (Fig. 7.17(c)).

Some optical switches are based on a tiny pivoting bar with a gold-plated micromirror at one end that fits in a tiny space between two optical fibers lined up end to end. In the Off state the micromirror rests below the cores (through which the light signal travels) of the two fibers, allowing light signals to travel across the gap from one core to the other. When a voltage applied to the far end of the bar helps to lift the mirror between the fibers, where it reflects incoming light rather than permits its passage.

The tortional type switch consists of tortional micromirror and corresponding electrode. The micromirrors and electrodes are fabricated on separate substrates as shown in Fig. 7.18. The electrodes

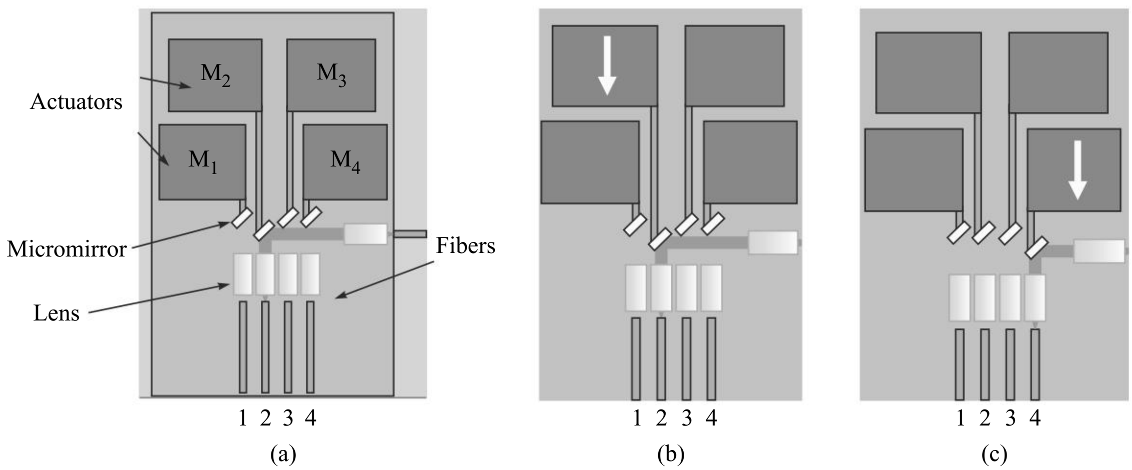


Fig. 7.17 *(a) 2D single input multiple output beam interception type optical switches (b) M_2 is activated by the electrostatic actuation (c) M_4 is activated by the electrostatic actuation*

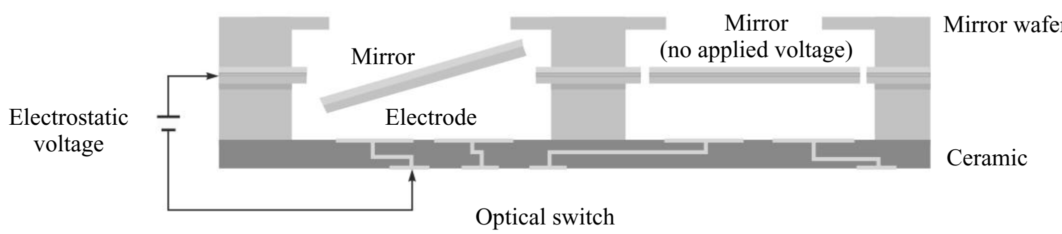


Fig. 7.18 *Schematic diagram of a micro optical switch with two tortional mirrors*

are situated on the ceramic layers. Special self-alignment mechanism are designed for assembling the mirrors and electrodes. If a bias voltage is applied to the electrode, the mirror is attracted by the electrode and tilted. The incident light thus can be reflected and redirected. The mirror is designed to tilt around x and y-axes.

7.12 WAVEGUIDE AND TUNING

A structure having the ability to guide light energy is called optical waveguide. A waveguide is called an optical conductor. The structure guides the light wave by restraining its travel along a defined optical path. The path is made up of dielectric material such as glass or plastic, called *core*. The design of the structure is such that the light is guided by following the principle of *total internal reflection* (TIR). The total internal reflection is possible only when the light is allowed to incident into the cross-sectional surface of the structure at an angle greater than the critical angle θ_c (Fig. 7.19(a)). For more information refer a book on Optical Communication Systems. However, a very brief description follows.

Light can be guided by planar, rectangular waveguides or by optical fibers as shown in the Fig. 7.19(b). The waveguide traps light along the optical path by its geometrical shape. The design consists of only two regions, the core and the surrounded cladding (Fig. 7.19(c)). The refractive index of the core, η_c, is greater than the refractive index of the cladding, η_{cl}, so that total internal reflection is achieved. The incident light is trapped as long as:

$$\sin\theta > \frac{\eta_{cl}}{\mu_c} \tag{7.11}$$

where θ is incident angle. Light is transmitted along the core and does not escape through the sides of the guide. A typical optical fiber based waveguide is shown in Fig. 7.19 (c), however, this is not a micro-optical waveguide. The outer part of this macro waveguide is called buffer, usually serves as a protecting layer for core and cladding.

In MEMS domain the dimension of an optical waveguide is extremely small; in the order of micrometer. A typical dimension of a microwaveguide is about 5 to 6 microns (1/10 the thickness of a human hair). These devices are therefore called micro-optical waveguides (MOW). The MOWs are also seen in the integrated optical chip (IOC). Much like electronic IC (here the discrete devices are diodes, transistors and resistors), an IOC could contain number of discrete devices such as splitter, coupler, modulator, attenuator, and switch connected to each other by MOW in a single substrate. Such an integrated optical circuit promises enhanced functionality, performance, compactness, and cost-effective volume production capability. A typical MOW shown in Fig. 7.19(d), is a tapered MOW and can typically be used for wavelength tuning and switching.

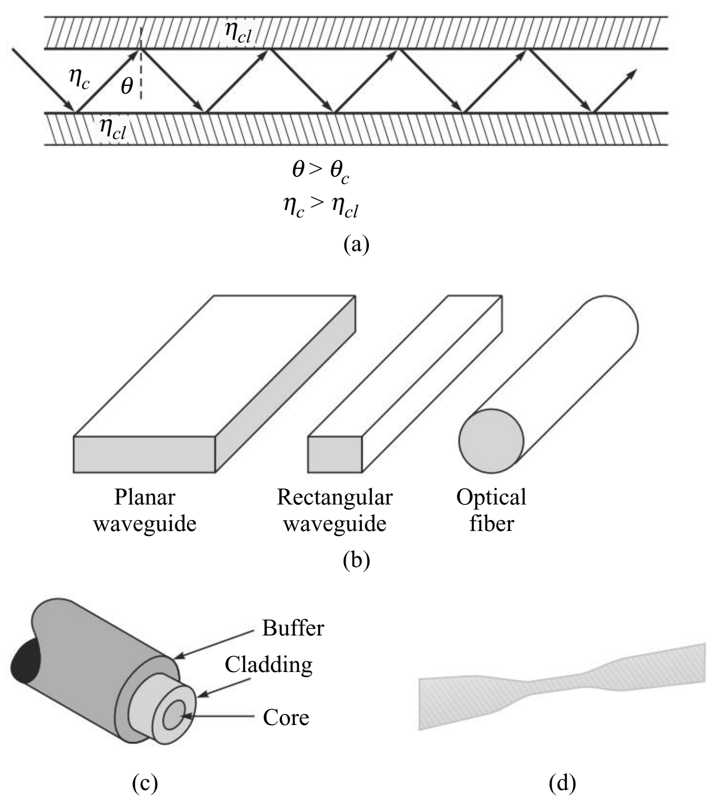


Fig. 7.19 *(a) The optical wave incidented into the structure at an angle greater than the critical angle; (b) Various types of optical waveguide; (c) A typical optical fiber showing core cladding and buffer (this is not a microwaveguide); (d) A tapered microwaveguide can be used for wavelength tuning, light switching (described in this section)*

7.12.1 Tuning

Tuning is defined as a method of selecting something. Wavelength tuning refers to the selection of a particular or group of wavelengths from a bunch. This also refers to demultiplexing, an opposite to multiplexing (MUX). A wavelength demultiplexer (DEMUX) separates the wavelength whereas a multiplexer combines the wavelengths. The fundamental behind the practice of tuning and demultiplexing remains same, but the application areas are somehow different although the terms are used synonymously.

Figure 7.20 illustrates an MOW based wavelength tuner cum switch. The wavelengths can be separated (hence tuned) and directed to a desired port or path. This design can be used for MUX/DEMUX applications. The system consists of two primary units; a tapered MOW and a microactuator. The MOW is placed on the microactuator platform (MAP) in between the upright projected strands and holding plates rigidly, as shown in Fig. 7.20(c). Out of the two strands one is movable. The MAP is also integrated with a microstep motor that can pull the movable strand over the rail attached to it at the bottom of the platform.

The tapered MOW is considered as a DEMUX or wavelength selective coupler. When it is used as DEMUX, the MOW shown in Fig. 7.20(a) can behave as a one-input and two-output port device. The input light composed of two wavelengths can be separated and directed through the two output ports (also called channels) by applying an axial stress (due to applied force) into the MOW. The axial stress is applied by the microactuator. The output port of this designated coupler thus can select a particular wavelength. The axial stress is exerted at the tapered portion of the MOW. The stress is built up because of the micro movement caused by the microactuator. The microactuator should have nanometer resolution range.

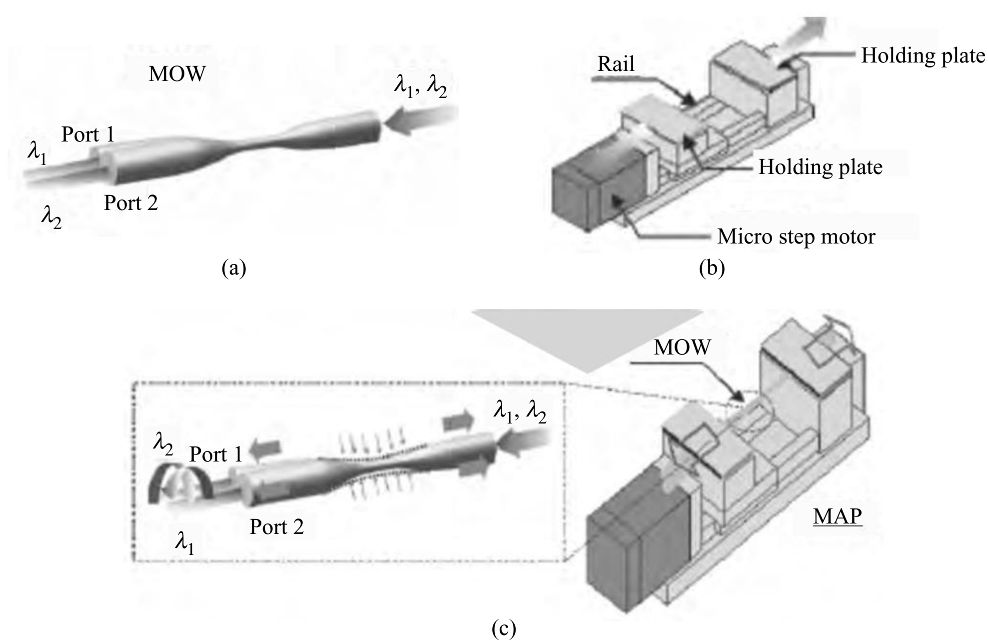


Fig. 7.20 *A tapered MOW; (b) The microactuator; (c) The tapered MOW integrated with the microactuator and subjected to axial stress*
Source: Shin, *Optic Express*, 12/19, 2004; *Courtesy:* Optical Society of America

The tapered coupler is characterized by a constant called coupling constant. The mechanical perturbation of the coupler (MOW) causes coupling constant to change. Mathematically the coupling constant is given by,

$$C(z) = \frac{\sqrt{\sigma}U^2(z)K_0(W)}{bV^3K_1^2(W)} \tag{7.12}$$

where $C(z)$ is called coupling coefficient, K_0 and K_1 are Bessel functions of second kind of 0^{th} and 1^{st} order respectively, U, V and W are normalized frequencies in a circular waveguide, the parameter z is axial displacement in z direction, b is the radius of the cylindrical waveguide in the coupling zone and σ is given by,

$$\sigma = \left(\frac{\eta_{cl}}{\eta_0}\right)^2 - 1 \tag{7.13}$$

where η_{cl} and η_0 are the refractive indices of cladding and air respectively. As can be seen from the above equation, the coupling coefficient depends on both dielectric constant and the dimension. Because of the applied axial force (stress) two effects are occurring. The photoelastic effect induces the refractive index to change and the stress-strain behavior changes the geometrical dimensions of the waveguide. As a result the coupling constant gets altered. The coupling constant in fact is a function of both the refractive index and the geometrical dimension. The change in coupling constant alters the power splitting ratio and spectral response at the output ports. Mathematically this can be understood as,

$$P_1(z) = \cos^2\left(\int_0^z C(z')dz'\right) \tag{7.14}$$

$$P_2(z) = \sin^2\left(\int_0^z C(z')dz'\right) \tag{7.15}$$

where, $P_1(z)$ and $P_2(z)$ are the power flow through the port 1 and 2, respectively, and as pointed out they depend on the coupling constant of the coupler (also DEMUX or waveguide).

Figure 7.21 shows a single-input four-output DEMUX. In this example, the input light signal contains four wavelengths such as λ_1, λ_2, λ_3 and λ_4 and needs to be separated. By employing similar method, i.e. axial stress method, the wavelength components can be separated and channeled through the desired output port as shown. In this typical case there should be three tapered couplers, which are to be stressed in two stages, stage-1 and stage-2. The 1^{st} stage coupler can split the input signal power into two components, by applying appropriate axial stress by the actuator. Now the Port-1 of the 1^{st} stage coupler may direct λ_1 and λ_3 components and Port-2 may direct λ_2, and λ_4 components, respectively. At the second stage the 2^{nd} and 3^{rd} coupler can separate λ_1, λ_2, λ_3 and λ_4 and direct them into four different ports. This DEMUX is known as a 4-channel WDM (Wavelength Division Meultiplexing) demultiplexer.

Further, by inducing phase shift at the taper zone, the wavelength selectivity can also be achieved. This suggests that the wavelength selection or separation can be achieved by two methods such as by varying the coupling coefficient and by inducing the phase shift. The permitted routing configuration is presented in Table 7.2. Here the first and the third columns are the phase shifts induced at the taper zone.

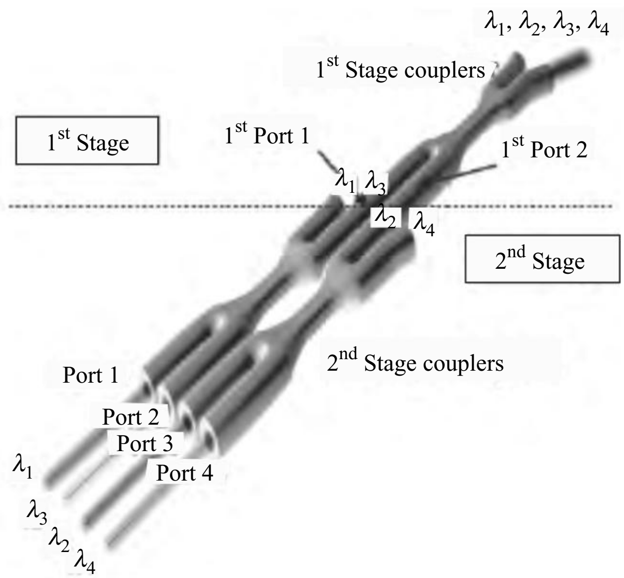


Fig. 7.21 *One-to-four DEMUX/Coupler/Tuner*
Source: Shin, *Optic Express,* 12/19, 2004; *Courtesy:* Optical Society of America

Table 7.2 *Permitted routing configuration*

1st stage	*2nd stage*		*Port* 1	*Port* 2	*Port* 3	*Port* 4
0	Upper	0	λ_1	λ_3	-	-
		$\pi/2$	$0.5\ \lambda_1 + 0.5\ \lambda_3$	$0.5\ \lambda_1 + 0.5\ \lambda_3$	-	-
		π	λ_3	λ_1	-	-
	Lower	0	-	-	λ_2	λ_4
		$\pi/2$	-	-	$0.5\ \lambda_2 + 0.5\ \lambda_4$	$0.5\ \lambda_2 + 0.5\ \lambda_4$
		π	-	-	λ_4	λ_2
π	Upper	0	$0.5\ \lambda_2 + 0.5\ \lambda_4$	$0.5\ \lambda_2 + 0.5\ \lambda_4$	-	-
		$\pi/2$	λ_2	λ_4	-	-
		$3\pi/2$	λ_4	λ_2	-	-
	Lower	0	-	-	$0.5\ \lambda_1 + 0.5\ \lambda_3$	$0.5\ \lambda_1 + 0.5\ \lambda_3$
		$\pi/2$	-	-	λ_1	λ_3
		$3\pi/2$	-	-	λ_3	λ_1

Columns 4 to 7 are the output wavelengths. Since the couplers split the power the configuration can also be used for power splitting applications.

7.13 SHEAR-STRESS MEASUREMENT

In many industrial and laboratory applications precise and accurate measurement of instantaneous wall-shear stress in a channel or pipe is required. Optical principles such as Doppler effect can be used for such precision measurement. Diverging fringe patterns are tested through Doppler shift to measure the gradient of the flow velocity at the wall. The principle of operation is described as follows.

Linear diverging interference fringes is projected from a surface and extended in the channel through which the fluid or gas is flowing (Fig. 7.11). The light will then be scattered because of the movable particles in the fluid. The frequency of the scattered light is proportional to the instantaneous velocity and inversely proportional to the fringe separation at the location of particle trajectory. The scattered light from the particle passing through the fringes is collected through a window integrated with a micro sensory system, i.e. a MOEM device. As particles in the fluid flow through the linearly diverging fringes, they scattered light with a frequency *f*. Then the wall shear can be measured as,

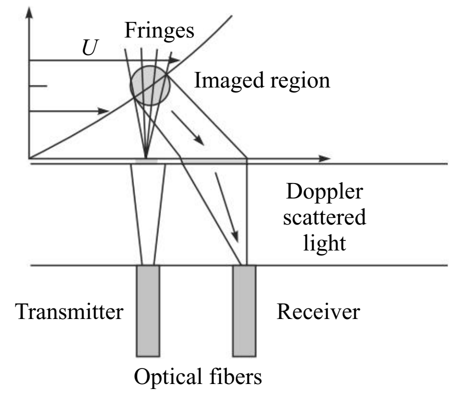


Fig. 7.22 *MOEM sensor for the measurement of wall shear*
Source: Fourguette

$$\sigma = \frac{\partial u}{\partial y} = \frac{u}{y} = k \times f \tag{7.16}$$

where, u is the velocity of fluid, k is called the fringe divergence rate, y is the distance from the sensor surface and f is the frequency of the scattered light. The above equation suggests that the Doppler frequency simply multiplied by the fringe divergence yields the velocity gradient.

The microsensor has two primary parts, the transmitter and the receiver. In practice, the light is passed through a single mode optical fiber and is allowed to diverge onto a Polymethyl-glutarimide diffractive lens. The light is filtered through two parallel slits of chromium layer. Then the fringed light is imaged onto a multimode fiber, which is coupled to a photodiode. The filter at the surface of the sensor ensures that the fringes originate at the surface of the sensor.

7.14 SUMMARY

In recent years, the impact of the microelectromechanical technology on our daily lives has probably become more profound than the impact of the microelectronic technology. Products fabricated with the emerging technologies of MEMS are being incorporated in an increasing number of sensor and actuator applications. Micromechanisms, to allow motion based on electrostatic actuation, when combined with electronics in order to manipulate the intensity and directionality of optical signal is called micro-optoelectromechanical systems (MOEMS). In other words, by combining microelectronics and micromechanical components, these devices promise to have a major impact on future sensors and actuators. Various thin films and related technologies such as the bulk and surface micromachining and LIGA technology for the elementary structures and devices of MOEMS have already been proposed. Moreover, different miniaturized optical analyzers employing MOEMS have been developed. There is rapid progress in MOEMS for telecommunications to meet the needs for increased bandwidth. Some applications require precise features for optical alignment while others involve the precise actuation of optical parts to achieve desired functionality. In the MOEMS approach accurate, low-loss, optical connections are made between different guided wave optical components including fibres and waveguides. To build three-dimensional MOEMS new materials have to be proposed.

In this respect this chapter deals with the important aspects of the MOEMS technology. The scope includes the study of properties of light and application domains of MOEMS. The principle, concept and design of optical switches, beam spliters and microlenses, DMD™ and GVL™ technologies are described with suitable schematic diagram and SEM photographs. The final section of the chapter discussed the principle of operation of a wavelength tuner using micro optic waveguide (MOW).

Points to Remember

- Opto-Electro-Mechanical (OEM) devices are claimed to be a hybrid type devices.
- When the OEM devices are designed at the microscale level they are called Micro-Opto-Electro-Mechanical devices or MOEM in short.
- Manufacturers of MEMS devices are forecasting new opportunities of MOEMS outside of the telecom industries including: information technology, healthcare, industrial test and measurement, military, and control sectors.
- Useful application specific MOEMS devices are microlens, micromirror, beam splitter, grating, waveguide and modulator.
- Reflection of light occurs when the light wave meets a surface or a second boundary that does not absorb the energy of the incident wave and bounces it from the surface.
- The speed of light varies from medium to medium.
- If light passes from one transparent medium to another, its speed changes, and as a result it does not travel in the same direction as it was traveling in the previous medium.
- As light is a wave, it has the capability to bend around corners.
- Interference is the superposition or overlapping of two or more waves resulting in a new wave pattern. Interference causes cancellation or amplification of the wave at that point.
- Light is considered as a mixture of *many* kinds of waves with *multitude of different wavelengths*. Each of these different waves vibrates in many different directions as the light travels.
- Light from moving objects will appear to have different wavelengths. In fact the wavelength of light shifts when a moving object emits it. This effect is called the Doppler shift.
- Light modulator modulates or changes some of the parameters of the light. The important parameters are wavelength and direction.
- Modulation of light is essential in high-end display applications.
- A beam splitter (BS) (Fig. 7.2) is a class of optical device that divides a light beam into two separate beams, hence the name beam splitter.
- Lens is used to focus the incident light beams.
- Two types of microlenses exist: Micro Fresnel lens and Micro spherical lens.
- Fresnel lens is flat on one side and ridged on the other side.
- MSL is also called micro-ball lens.
- Micromirrors are circular, rectangular or square in structures.
- Optical switching through micromirror is achieved by integrated actuation mechanism.
- Micromirrors are produced as single element containing a single micromirror or a group of micromirrors in a single wafer in a two dimensional array.
- DMD™ is a micromechanical silicon chip. The DMD™, a Spatial Light Modulator (SLM™), is primarily used for imaging applications.
- The DMD™ is basically considered as an ultra precision light-switching device that is capable of modulating the light digitally by the use of huge numbers of microscopic mirrors, which have been arranged in a definitive manner, usually in the form of rectangular array. The DMD chip contains movable aluminium mirrors integrated with electronics, logic, memory and control circuitry.

- A diffraction grating (DG) is a set of parallel slits (equivalent to obstacles that are placed alternatively) used to disperse or bend the light. DG can be used to separate the dense wavebands.
- Another spatial light modulator technique characterized itself as Grating Light Valve™ (GLV™) has taken momentum in image display applications. Unlike DMD™ technology, GLV™ technology works on the principle of diffraction.
- The GLV™ device consists of parallel rows of tensile reflective membrane type micro-ribbons.
- Applying electrical voltages can pull the ribbons vertically down relative to the underlying surface.
- The GLV™ device is essentially an addressable dynamic diffraction grating, due to the reason that the ribbons in a pixel are controlled in terms of addressing them by the use of electronic circuitry.
- The micromirrors can route light signal from and to any input/output optical fibers of 2D or 3D array. So they can be called switches. There are mainly two types of switches based on the way the mirrors are activated. They are interception type and rotation type. Irrespective of their types the optical switches are activated using the principle of electrostatic actuation.
- A structure having the ability to guide light energy is called optical waveguide.
- The construction of the optical waveguide is such that the light is guided by following the principle of *total internal reflection*.
- Tuning is defined as a method of selecting something. Wavelength tuning refers to the selection a particular or group of wavelengths from a bunch.
- A demultiplexer (DEMUX) separates the wavelength whereas a multiplexer combines the wavelengths.
- In many industrial and laboratory applications precise and accurate measurement of instantaneous wall-shear stress in a channel or pipe is required. Optical principles such as Doppler effect can be used for such precision measurement.

Exercises

1. Expand the abbreviation MOEMS.
2. How is a hybrid system expressed?
3. What are the important applications of MOEMS devices?
4. What are the relative merits of MEMS based optical devices?
5. Review the properties of light and define the following.
   (a) Reflection (b) Refraction
   (c) Interference (d) Polarization
   (e) Doppler shift
6. What is the function of a typical light modulator?
7. Is beam splitter a light modulator?
8. How many types of MEMS microlens do you know? Explain their design features.
9. Discuss in detail how a spherical microlens is fabricated.
10. Write notes on MEMS micromirror.
11. Clearly discuss the principle of operation of the following MEMS products.
    (a) DMD (b) GLV
12. Explain the principle of operation of various types of optical switches used in optical communication systems.
13. What is a waveguide? Show different types of waveguides. How is tuning achieved using the micro-optical waveguide (MOW)? Define MUD and DEMUX.
14. Explain the principle of measurement of shear stress by using MOEMS devices.

Chapter 8

Magnetic Sensors and Actuators

Objectives

The objective of this chapter is to study the following.

- Magnetic MEMS (MagMEMS) in general
- Magnetic materials and their properties
- Magnetization, Magnetoresistivity and Hall effect
- Magnetic sensor fundamentals
- Magnetodiode and Magnetotransistor
- Magnetic actuator fundamentals
- Magnetic pressure sensor using MOKE
- Magnetic actuation based RF switches
- Bi-directional actuator fundamentals, operation and fabrication
- Magnetic actuator for microsurgery applications
- Large-force integrated VR actuator
- Fundamentals to magnetic probe based storage devices

8.1 INTRODUCTION

MEMS technology has the greatest impact on sensors and actuators. The fabrication of miniature devices such as microsensors and microactuators has long been the main thrust of micromachining technology. MEMS is an expansion and evolution of IC based technology. The possibilities with MEMS technology involve the utilization of a range of materials and processes not only to obtain precise dimensional scale but also to improve operational flexibility and to reduce mechanical complexity. Much of MEMS technology is silicon based, with 3D structures being fabricated from a silicon platform using various lithographic processes. Magnetic properties of materials continue to drive the development of new technology. Micromachining technology has already began to develop magnetic MEMS, called MagMEMS. MagMEMS includes sensors, actuators and storage devices. MagMEMS are being developed using MagMEMS materials for a variety of purposes.

The important applications of magnetic sensors are to measure,

- Local and uniform magnetic field (magnetic compassing)
- Movement and direction detection of large objects (Car, Bus, Train, etc.)
- Rotation (Rotating shaft, devices, etc.)
- Current density and direction (Electrical machines)
- Linear position (Proximity sensing)
- Movement (Head tracking for virtual reality equipment)

The applications of magnetic actuators are many including microsurgery, optical switching and data writing and reading. Magnetic actuation is an attractive driving principle and very relevant for dust-filled environment, conducting fluid and in an environment where low driving voltages are acceptable and/or desired.

Regardless of their applications, this chapter presents the fundamental principle of operation of important types of MagMEMS devices. In particular magnetodevice, RF switch, pressure sensor, probe-based storage device and bi-directional actuator are the focus of study.

8.2 MAGNETIC MATERIALS FOR MEMS AND PROPERTIES

MagMEMS materials are primarily suitable for many MEMS applications and offer several advantages over other materials. The suitable materials are those having responsive characteristics like high inductive coupling, permeability and resonant frequency. In many occasions these characteristics are exploited to map with the required sensory or actuating behavior. Not only magnetic materials are used for sensors and actuators but can be used for other purposes. For instance, they are used in MEMS device that may require a deposition of magnetic thin film as in case of storage sled (see Section 8.14). The important issues that are to be addressed in MagMEMS materials are, integration and compatibility with CMOS[1] technology, stress compensation, diffusion process and self-tuning behavior. Magnetic films have been successfully deposited on a wide range of substrates such as Si or GaAs, polymers and glass.

Broadly, MagMEMS materials are classified into three categories, such as,

- Magnetoresistive materials
- Magnetostrictive or Magnetoelastic materials
- Permanent (hard) magnetic materials

8.2.1 Magnetoresistive Materials

A technologically relevant magnetic phenomenon is magnetoresistance (MR), in which the electrical transport properties are strongly correlated to the applied magnetic field. That is the resistivity of magnetoresistive material changes as the function of the external magnetic field. The magnetic field disturbs or modulates the scattering of charge carriers such as electrons in the material. The effect is exploited for magnetic MEMS devices.

8.2.2 Magnetostrictive Materials

There exist some materials which when placed near a magnet lend to change their shape. The effect is called magnetostriction, hence the name magnetostrictive material. Magnetostrictive materials are also

[1] Note that some integrated MEMS devices incorporate CMOS technology.

called active as well as soft magnetic materials. An example of Magnetostrictive material is Terfenol-D. An internal magnetoelastic stress results, in a bulk of material, when the direction of magnetization in a magnetic material is rotated. The magnetoelastic stress gives rise to a magneotstrictive strain, which indeed is defined in terms of magnetostriction. The magnetostriction is molecular in origin for which the mechanical response is very fast. Magnetostrictive materials have a coupling effect like piezoelectric material, which means that a change in one state produces a change in the other. Magnetic field produces shape change and therefore, can be used advantageously as actuators, while stress change produce magnetic change, thus can be used as sensors. Practical devices normally require large magnetostriction and a small magnetic anisotropy. Large magnetostriction signifies large range of motion. The power produced by the actuator depends on the square of the magnetostriction. Therefore, small improvements in the magnetostriction can produce large output. A small magnetic anisotropy implies large change in length in response to a small magnetic field. While magnetostriction refers operational range of motion anisotropy broadly signifies sensitivity. Advances in the processing of terbium-dysprosium have yielded a unique polycrystalline structure in the material that significantly increases magnetostrictive effect. The magnetostrictive property of the materials is temperature dependent. Curie Temperature (CT) is an important property of magnetic material. CT is the temperature at which a material loses its magnetic properties. The magnetic property disappears as a result of thermal agitation. CT is normally high but lower values are always desired in order to operate the device in the ambient temperature. Typical desirable ranges of operations are from –280 to 600 °C. Alloys of cobalt ferrite with silicon and manganese are being developed to reduce the Curie temperature to room temperature. The new class of materials is expected to significantly improve magnetoelastic characteristics over competitive materials. Table 8.1 illustrates the CT (in °K) of materials.

Table 8.1 *Curie temperature of several materials*

Material	*Curie temperature (in degree K)*
Co	1387
Fe	1043
$FeOFe_2O_3$	858
$NiOFe_2O_3$	857
$CuOFe_2O_3$	728
$MgOFe_2O_3$	712
MnBi	631
Ni	627
MnSb	587
$MnOFe_2O_3$	571
$Y_3Fe_5O_{12}$	561
CrO_2	386
MnAs	318
Gd	292.7
Dy	89
EuO	69.5

8.2.3 Hard Magnetic Materials

Most materials we encounter have no magnetic properties. In these materials, the magnetic fields of the individual atoms are randomly aligned so that the net magnetic effect is zero. In a permanent magnet, however, the magnetic fields of the individual atom are aligned in one preferred direction, giving rise to a net magnetic field. Permanent magnetic materials are called hard magnetic materials. The permanent magnets have high magnetization, which allows the use of such magnets having the extreme shapes and small dimensions as required in devices.

The relationship between the magnetic filed and magnetic flux density B in free space is given by,

$$B = \mu_0 \bullet H \tag{8.1}$$

where μ_0 is the permeability of the free space and is given by, $4\pi \cdot 10^{-7}$ V·s/A·m.

When one demonstrates the effects within the magnetic material, then the relationship between the magnetic fields and the flux density is given by,

$$B = \mu_0 (H + M) \tag{8.2a}$$

where M is called magnetization of the magnetic material. Magnetization is understood as the magnetic dipole moment per unit mass and is measured in amperes per meter. The applied field is also measured in amperes per meter.

An alternative form of the above equation is

$$B = \mu_0 \bullet \mu_r \bullet H \tag{8.2b}$$

where μ_r is called relative permittivity of the material. From the above equations, it follows that:

$$\frac{M}{H} = (\mu_r - 1) \tag{8.2c}$$

$$\Rightarrow \qquad \chi = (\mu_r - 1)$$

where χ is called magnetic susceptibility of the material. The magnetic materials are classified according to their magnetic susceptibility and permeability. Table 8.2 shows classification of some important magnetic materials. Ferromagnetic materials are mostly used for the design of MagMEMS devices as they have high value of relative permeability. Cobalt, iron, nickel, and some of the rare earths such as gadolinium and dysprosium exhibit ferromagnetism.

Figure 8.1 illustrates the behavior of a ferromagnetic material when they are magnetized, demagnetized and re-magnetized. They show hysteretic properties. In the plot M_s, M_r and M_c are defined as the *saturation magnetization*, *remanent* magnetization and *coercivity*. H_s is called saturating magnetic field. Soft magnetic materials have small coercivity and low saturation field whereas hard magnetic materials have high values.

Table 8.2 *Classification of magnetic materials*

Category	χ/μ_0	*Examples*
Ferromagnetic	10^7 to 10^2	Ni, Fe, Co, NiFe, NdFeB
Ferrimagnetic	10^4 to 10^1	Fe_3O_4, ferrites, garnets
Antiferromagnetic	small	MnO, NiO, $FeCO_3$
Paramagnetic	10^{-3} to 10^{-6}	Al, Cr, Mn, Pt, Ta, Ti, W
Diamagnetic	-10^{-6} to -10^{-3}	Ag, Au, C, H, Cu, Si, Zn
Superconducting	-1	$YbBa_2Cu_3O_x$

Source: Judy et al., U of California

Example 8.1 Express the relationship between the magnetic susceptibility and absolute permeability. Discuss the meaning of diamagnetic and paramagnetic materials from the expression of relative permeability. Define mass magnetic susceptibility and molar magnetic susceptibility.

Solution: Another way of expressing the relationship between the magnetic susceptibility and absolute permeability, μ can be written as,

$$\mu = \mu_0(1 + \chi)$$

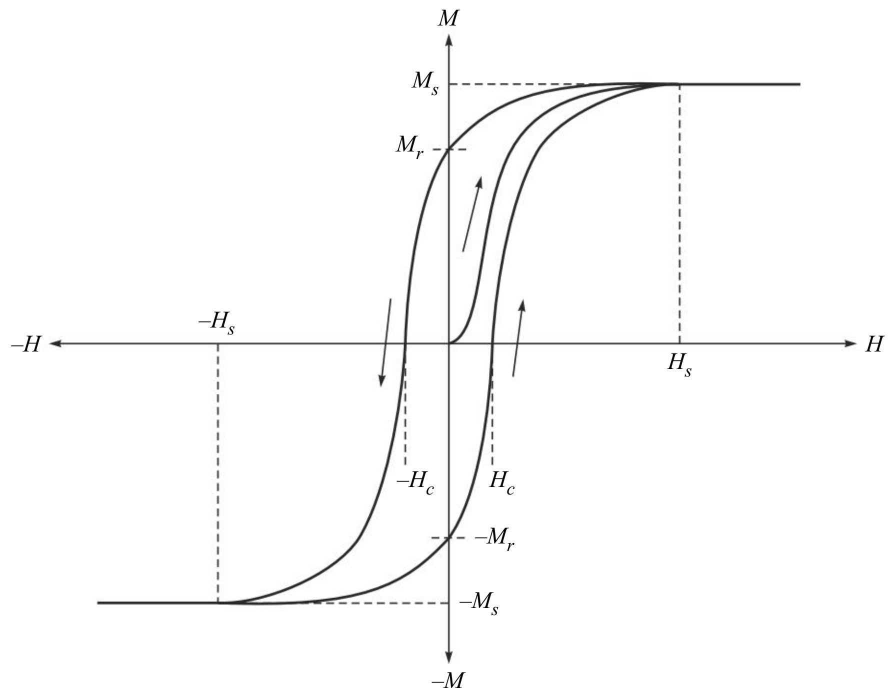


Fig. 8.1 *A plot of magnetization as a function of magnetic field (H_s = Saturation magnetic field; $-H_s$ = Opposite saturation magnetic field; M_s = Saturation magnetization; M_r = Magnetization retentivity; $-H_c$ = Opposite magnetic coercivity; $-M_s$ = Opposite saturation magnetization; $-M_r$ = Opposite magnetization retentivity; H_c = Magnetic coercivity*

As obvious, $(1 + \chi)$ is the relative permeability of the material. In order to obtain *SI* value of susceptibility, the cgs (centimeter-gram-second) value of susceptibility is multiplied by 4π. Further, if χ is positive, i.e. for $(1 + \chi) > 1$ the material is called paramagnetic. On the other hand, if χ is negative, $(1 + \chi) < 1$ the material is known as diamagnetic. In the former case, the magnetic field is strengthened by the presence of the material. With regard to measurement of susceptibility, there exists two types, one is called mass magnetic susceptibility which is expressed in cm^3g^{-1} (χ_g) and the second one is called molar magnetic susceptibility (χ_M) that is expressd in cm^3mol^{-1}. Mathematically they are expressed as

$$\chi_g = \chi/\rho$$

$$\chi_m = M\chi_g = \chi M_m/\rho$$

where ρ is the density in gcm^{-3} and M_m is molar mass in $gmol^{-1}$.

8.2.4 Design Considerations in Magnetic Materials

At the time of fabrication high values of stress is developed in the MagMEMS structures. This is undesirable. Inner diffusion between the magnetic material layer and the immediate layer may occur causing performance degradation of magetic film. Care must be taken in this respect. Magnetic materials are susceptible to external interference. Self-induction and external biasing should be taken into account while designing the MagMEMS devices. The deposition of stress compensation layer needs to be considered.

8.3 MAGNETIC SENSING AND DETECTION

Magnetic sensors have long been in use. Primitive applications were for direction finding used in navigation. Although, magnetic sensors are still a primary means of navigation, many more uses have evolved so far. The technology for sensing magnetic fields has also evolved driven by the need for improved sensitivity, capability and smaller size with in-built signal conditioning electronic systems. The way magnetic sensor differ from the other types of sensors is illustrated in Fig. 8.2.

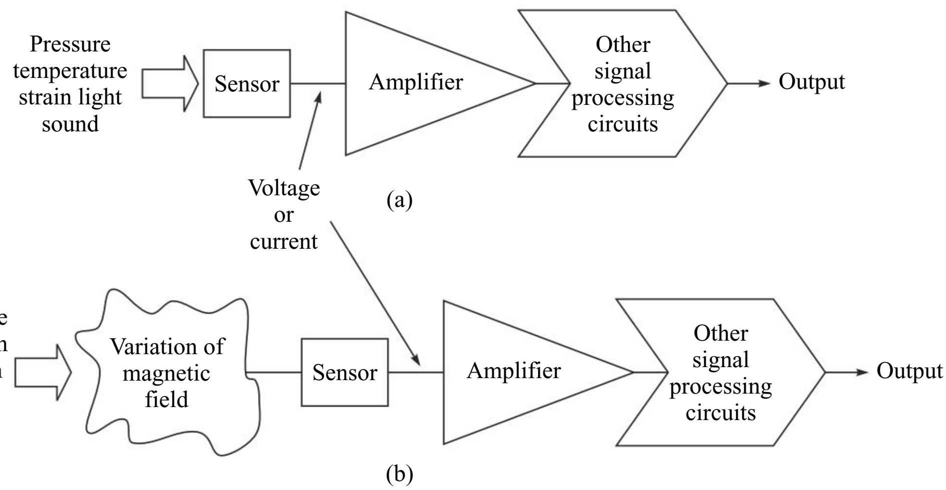


Fig. 8.2 *(a) A typical sensor, (b) An illustration of magnetic sensor*
Courtesy: Honeywell, Inc.

8.3.1 Presence and Direction Detection of Large Object—An Example

Magnetic sensors perform extraordinarily in the magnetic field of below 1 gauss. The sensor can help in detecting ferrous objects which disturb the magnetic field of the earth. The presence, direction of movement and classification of objects such as airplanes, trains, automobiles, etc. can be determined by the use of these sensors. Since the earth's field provides a uniform magnetic field (UMF) over a wide area, the movement of vehicles can disturb the UMF, as shown in Fig. 8.3. Anisotropic Magnetoresistance (AMR) magnetic sensors in this respect are good in detecting the UMF disturbance caused by the vehicle. The field disturbance features can be mapped to classify different types of vehicles: cars, vans, trucks, buses, trailer trucks, and so on. The disturbance features corresponding to different large ferrous objects are modeled and stored as a composition of many dipole magnets. In principle these dipoles constituting the object have north–south and hence orientations that cause distortions in the magnetic field.

8.4 MAGNETORESISTIVE SENSOR

A very simple passive example is illustrated in Fig. 8.4. An AC current sets up time-varying magnetic field whose gradient exerts force on a high-permeability magnetic material placed at the end of an MEMS cantilever resonator, which vibrates and generates piezoelectric output voltage. The voltage can be measured in order to measure the current strength.

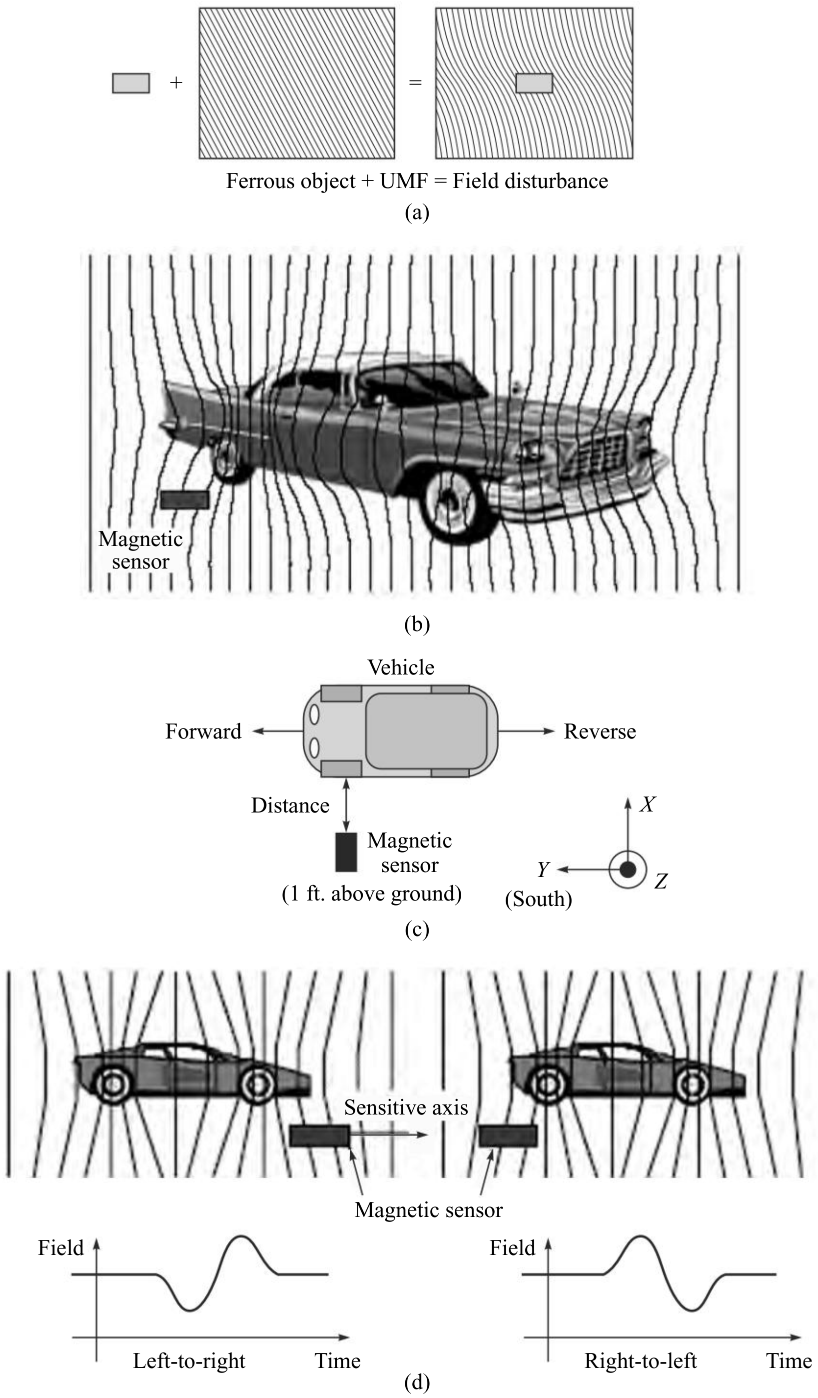


Fig. 8.3 *(a) Ferrous object disturbing UMF, (b) A typical example with car, (c) A schematic to represent direction sensing, (d) Direction sensing for vehicles*
Courtesy: Honeywell Inc.; *Source:* Caruso.

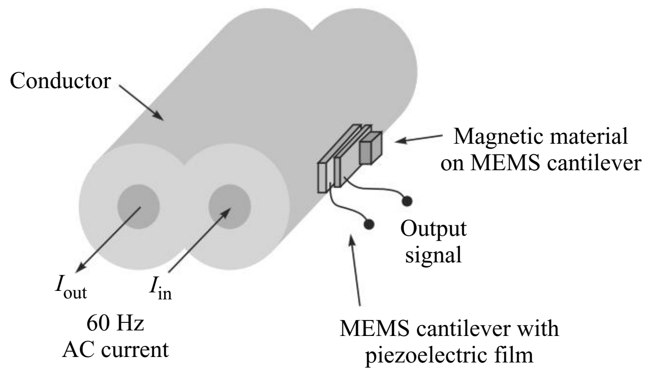


Fig. 8.4 *An MEMS cantilever based magnetic sensor*

8.4.1 Principle

The magnetoresistive effect has increased in practical importance in the last thirty years, owing to the advanced technology. Magnetoresistive (MR) sensors make use of the magnetoresistive effect, the property of a current-carrying magnetic material to change its resistivity in the presence of an external magnetic field. The magnetoresistive effect was described in 1857 by W. Thomson, who found that a magnetic field applied to a ferromagnetic material caused its resistivity to change. The amount of change depends on the magnitude of the magnetization and the direction in which the current is flowing. Magnetoresistive sensors are the most sensitive solid-state magnetic devices available to measure direction and magnitude of magnetic fields typically ranging from 70 mG to 6 G. Magnetoresistive sensors can also determine the change in earth's magnetic field due to the presence of a ferromagnetic object. There are primarily two physical effects where the resistance is dependent on the presence of magnetic field. These are called

- Anisotropic Magnetoresistive (AMR) effect
- Hall effect

Anisotropic magnetoresistive effect is associated with the change of the internal magnetic domains direction in the material. Further, clarification states that the *specific resistance* of the material in one direction with the internal magnetization vector is little higher than the *nominal resistance* in the orthogonal direction. Employing appropriate material processing method one can obtain the linear dependence of the change of resistance. During the fabrication of the device (sensor) a magnetic field M, is applied along the length of the strip in order to magnetize it. The direction of this magnetic field is referred to as easy-axis or anisotropic-axis direction. In real situation the device is equipped in such a manner that the current should pass in a direction that is 45^{o} with respect to the anisotropic axis (Fig. 8.5). Then a magnetic field H, is applied to the magnetization vector M, causing the magnetization vector to rotate. This in turn changes the magnetoresistance of the strip. The change in resistance is about 2 per cent. The mathematical description to this effect is given by,

$$\rho(\theta) = \rho_1 + (\rho_2 - \rho_1)\cos^2\theta \tag{8.3}$$

where, ρ_1 is the resistivity for angle $\theta = 90$ degrees and $\rho_2 - \rho_1$ is called anisotrophic resistivity change.

The improvement of the technology of thin ferromagnetic films (with a thickness of 10–50 nm) and the utilization of the Anisotropic (Fig. 8.5) Magnetoresistive (AMR) effect led to an increasing technical interest. A class of magnetoresistors with larger sensitivity than the standard MR devices is known as Giant Magneto Resistors (GMR). These materials are used as virtual reality position sensors and also as

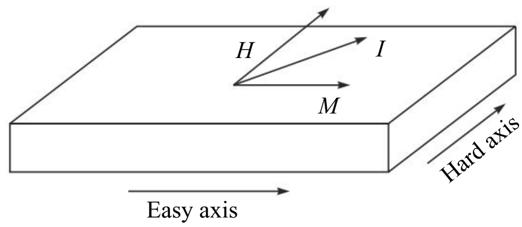


Fig. 8.5 *A schematic illustration of anisotropic thin film magnetoresistor*

hard disk drive read/write heads. The maximum resistivity change is up to 80% with the GMR effect, albeit at very high magnetic fields.

Hall effect based on the Lorentz force states that the resistance of the material is mostly dependent on the magnetic flux density B as presented in the following systems.

8.5 MORE ON HALL EFFECT

The principle of operation of Hall effect transducers is as follows. If a metal carrying a current is placed near the magnetic field in a proper manner, the charge carriers in the metal experience a force. As a result of this, the charge carriers are forced to get displaced. The phenomenon is called *Hall effect*, which was originally discovered by E.R Hall in 1879. The observation is illustrated in Fig. 8.6. The direction of the displacement depends on the direction of current and the magnetic field. For optimum effect essentially, the direction of the current and the direction of the magnetic field is kept perpendicular to each other. Under this condition, the displacement of electrons will take place towards the lower surface of the metal bar and a same amount of positive charge will appear on the upper surface. The accumulated charge carriers create an electrostatic field or potential difference across the edges at right angles to the current flow. The voltage so produced is called Hall voltage, V_h. Mathematically, the electrostatic field, ξ so developed can be expressed as,

$$\xi = v_d B \tag{8.4}$$

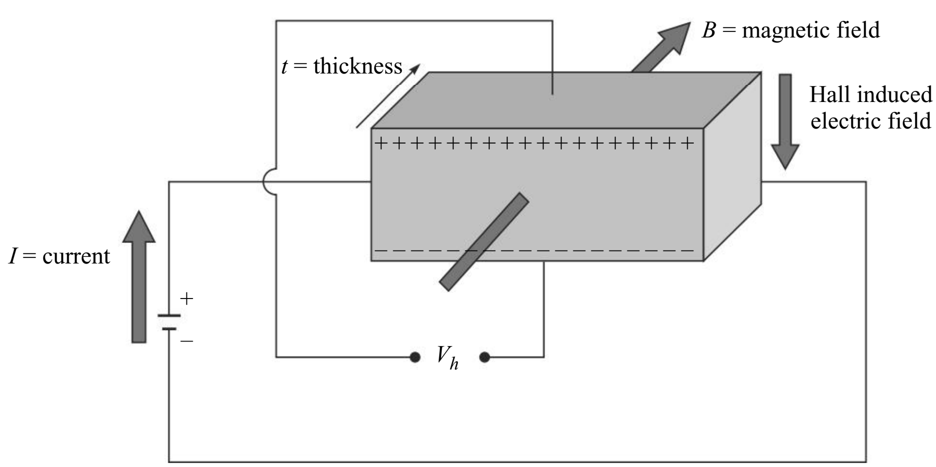


Fig. 8.6 *Hall effect*

In which,

$$v_d = J/nq = I/nqA.$$

v_d is called the drift velocity
B is the magnetic field
q is the charge of electron (1.602×10^{-19} Coulomb)
n is the number of electrons displaced
J is the current density
A is the cross-sectional area of the metal along the vertical axis
I is the current.

The relationship between the electric field and the Hall voltage can also be written as,

$$V_h = \xi h \tag{8.5}$$

where, h is the height of the metal bar. From Eq. 8.4 and Eq. 8.5, we have,

$$V_h = v_d Bh = \frac{IBh}{nqA} = H_c \frac{IB}{t} \tag{8.6}$$

$H_c = 1/nq$ is called Hall effect coefficient. The coefficient is negative for n-type semiconductor; see Table 8.3. $t = A/h$ is the thickness of the bar.

From Eq. 8.6, one can observe that the Hall voltage is proportional to the current and the magnetic field. If the current is deliberately made constant (usually 10–50 mA) the Hall voltage is proportional to the magnetic field. Either the current or the magnetic field can be measured by using Hall effect transducers.

Table 8.3 *Hall coefficient of some materials*

n-type material	*Hall coefficient*
Bismuth	–5*10E-7
Copper	–5.33*10E-11
Germanium	–3.5*10E-2
Indium antimonide	–6*10E-4
Silicon	–1.0*10E-2

Hall effect is pragmatic in all metals, but for the fabrication of the transducers, n-type semiconductor materials are preferred because it is easier to dope a controllable amount of charge carriers into the semiconductor material depending on the application domain of the transducer.

Typically, Hall voltage is a very weak signal, in the order of few microvolts for a magnetic field of one gauss. Therefore, proper electronic circuits are integrated with the transducer. The magnetic field could either be positive or negative, if it is employed for the measurement of current. Hall sensors are used in automotive, automation and control, robotics, valves, computers, motor commutation, security systems where essentially direction sensing, position detection, flow-rate detection, angular movement, length measurement and current sensing are desired.

As an example, in electrical machines, faults can be detected by measuring the currents or voltages in the machine windings. The magnetic field created near electrical machines changes when faults occur and therefore using such transducers to measure flux can aid in condition monitoring.

Example 8.2 Calculate the hall voltage that is developed if the applied magnetic field is 0.70 tesla and supplied current is 100 mA to a thin copper plate with cross-sectional area 0.75 mm^2 with thickness 2 mm.

Solution: Given: The applied current, I = 100 mA = 0.1 Amp.

Magnetic field, B = 0.7 tesla

Area = 0.75 mm^2

Thickness = 2 mm

Now height, $h = A/t = 0.75/2 = .375 \text{ mm} = 375 \times 10^{-6}$ m.

Mass of copper = 63.54 g/mol = 63.54×10^{-3} kg/mol

Density of copper = 9 g/cm^3 = 9×10^3 kg/m^3

Hence the number of free electrons per mol is same as that of Avogadro's number, which is 6.02×10^{23}/mol

So, the number of free electrons per unit volume will be,

$$n = \frac{density \times Avogadro's\ number}{Molar\ mass} = 8.5 \times 10^{28} \text{ electrons/m}^3$$

Now, the Hall voltage,

$$V_{th} = \frac{IBh}{nqA} = \frac{IB}{nqd} = 2.5735 \times 10^{-9} \text{ volt.}$$

Charge of electron is $= 1.60217646 \times 10^{-19}$ coulombs

8.6 MAGNETODIODES

Magnetodiodes and magnetotransistors are two semiconductor-based sensory devices used for the detection of magnetic field. The magnetodiodes use the effect of magnetoconcentration property. Magnetoconcentration is the concentration gradient of the carrier perpendicular to the magnetic vector and the current direction. The magnetodiode is made up of p-i-n semiconductor diode. The magnetodiode is very small in size and fabricated on a thin Si film of SOI (Silicon on Insulator) substrate. It is basically a *pn* junction, where *p* region is separated from the *n* region by an area of undoped silicon. Arguably, magnetodiode is not considered as MEMS product although it is a microdevice. The magnetodiode is roughly three times more sensitive than the Hall effect sensor. Magnetodiode inherits fine resolution capability for which it is used to measure local magnetic field in a positioning and fault detection application.

SOI (Silicon-on-Insulator), SOS (silicon on sapphire; Al_2O_3 is called sapphire) and MOS (Metal-Oxide-Semiconductor) are commonly used technology for fabrication. Out of these an appropriate method is adopted considering the area of application of the device. These days several sensors belonging to the magnetodiodes family are realized with either SOI or SOS technology, taking into account of power consumption, integration flexibility and input impedances.

As shown in Fig. 8.7, the recombination surface is formed by Si (or Ge) film deposition on the top surface of p-region (i.e. high resistivity region) of the p^+-p-n^+ diode structure. Sputtering deposition of thin film of Si followed by annealing forms the recombination layer. Two views such as upper and cross-sectional view of the magnetodiode formed on a SOI p-Si(100) substrate are shown in the Fig. 8.7(a). p-Si(100) is the SOI layer which has resistivity ranging from 20–30 Ω cm. The thickness is approximately 2 mm and the orientation is about 100. The process of chemical etching removes the SOI layer on the surface except the pattern of two electrodes and the recombination region. The dimensions of the high resistivity reason are about length L, ~100 μm and width W, ~40 μm. The operational principle of this magnetodiode is as follows. The operation is based on deflection of injected carriers by Lorentz force to the recombination surface side (Fig. 8.7(a)). The carriers are injected due to the external biasing voltage.

Assume that the magnetodiode is forward bias. The electrons and holes are produced by double injection in the forward biased p-p-n magnetodiode. They are deflected towards the recombination

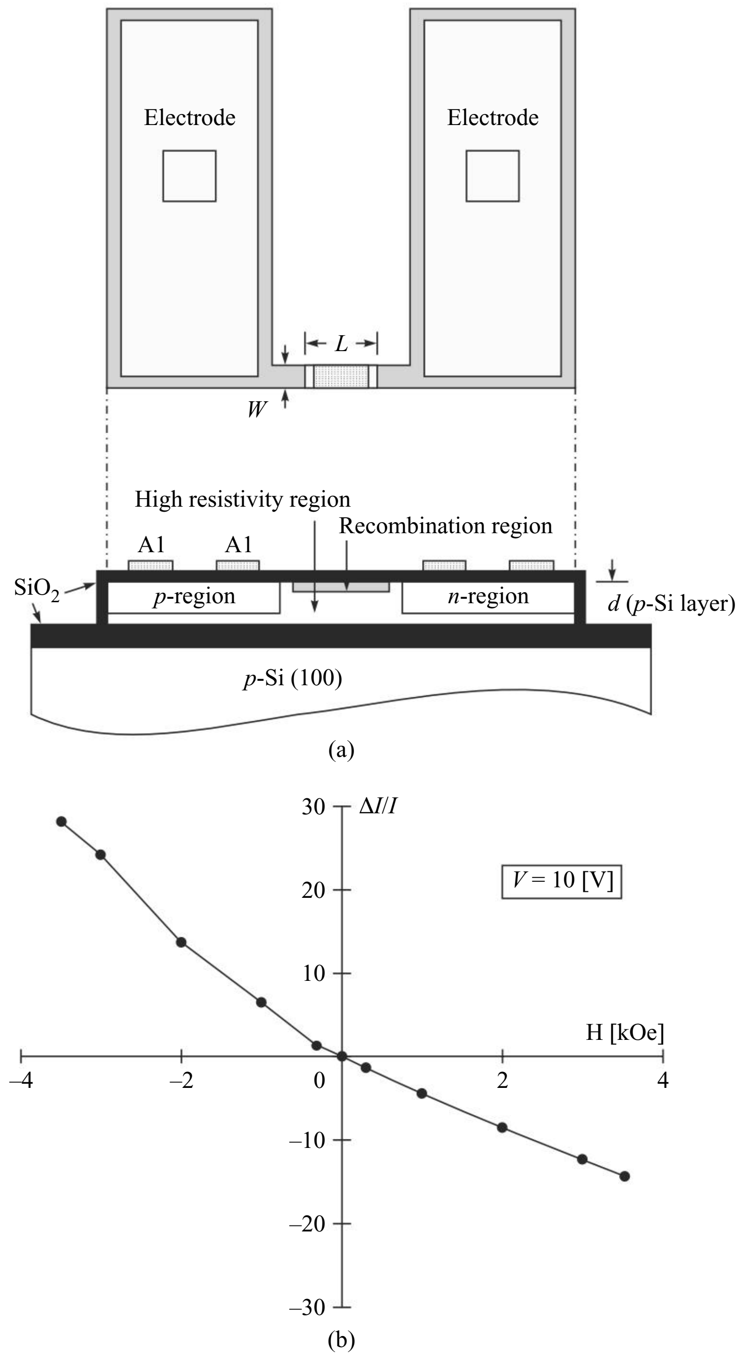


Fig. 8.7 *(a) Upper and cross-sectional view of the magnetodiode formed on a SOI p-Si(100) substrate, (b) Magnetic sensitivity*
Source: Hiroshi et al.

region by Lorentz force when magnetic field H is applied. Once the deflected carriers available in the recombination region are recombined they disappear. The current contributed by these deflected carriers is called recombination current, ΔI. This in turn causes the forward current I of the magnetodiode to decrease. On the other hand the forward current I is increased for reverse magnetic field H, because of the absence of recombination region in the other side. The magnetic sensitivity of magnetodiode is defined as the ratio of recombination current and the applied magnetic field at a given biasing voltage.

In Fig. 8.7(b) characteristic curve of magnetic sensitivity at biasing voltage of 10 volts is shown. The characteristic curve is a plot of a fraction of forward current $\Delta I/I$, expressed in percentage, versus applied magnetic field H. The curve is apparently linear, but as can be observed the ratio is relatively large corresponding to reverse magnetic field. $\Delta I/I$ is high at reverse magnetic field of 3.5 kOe. The sensitivity of the magnetodiode depends on the constructional design and on the passive parameters such as magnetoresistance, which in turn depends on the size of the depletion layer (space charge region), and on the contact potential at the interface between two semiconductors with different Fermi levels. Also the sensitivity depends on the magnetic field strength. In a magnetodiode the relationship between the forward current I, applied voltage V, and the depletion layer L, is as follows.

$$I \alpha \frac{V^2}{L^3} \tag{8.7}$$

The SOS technology is a recent development. SOS-based magnetodiodes are p^+-n-n^+, Schottky and filamentary magnetodiodes. SOS devices are drawing a lot of attention, as they are capable of performing high-speed operations while consuming low amounts of power.

8.7 MAGNETOTRANSISTOR

The magnetotransistor possesses very large sensitivity compared to magnetodiode. The magnetotransistor shown in the Fig. 8.8 is an npn transistor with two collector terminals, two base terminals and one emitter terminal. The emitter region and collector regions are heavily doped and both are n^+-type. The base is p^+-type, resides in between the emitter and collector and also heavily doped. The transistor has a recombination region close to the base region.

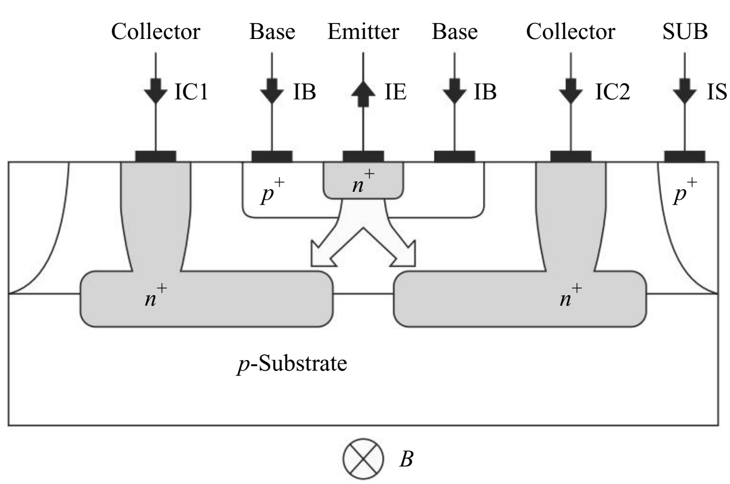


Fig. 8.8 *Schematic of Magnetotransistor*

While in operation the carriers are injected because of forward biasing of emitter-base junction. The operation of magnetotransistor is based on collector current detection. The collector current of the magnetotransistor is in fact modulated by the presence of magnetic field. The modulation is rather called emitter injection modulation, as in a typical transistor the collector current is proportional to the emitter current. The emitter injection modulation occurs due to the induced Hall field in the emitter region. When injected carrier such as electrons, in a typical configuration, coming from the emitter region penetrated into the base region, are deflected toward the recombination region (in the vicinity of base region) due to the presence of external magnetic field (to be measured), the conductivity of the injected carrier is drastically reduced, thereby resulting in a reduction in collector current. Because of deflection the electrons come upon longer traveling path in the base region and therefore get sufficient time to recombine with the majority carriers, the holes. The amount of deflection is proportional to the magnetic field strength. If most injected electrons recombine with majority carrier holes in the base region, collector current will be almost zero. This implies that it has large sensitivity. The magnetotransistor based sensor is fabricated in standard bipolar IC technology. The unijunction, bipolar, and field-effect transistors are the various forms of magnetotransistors.

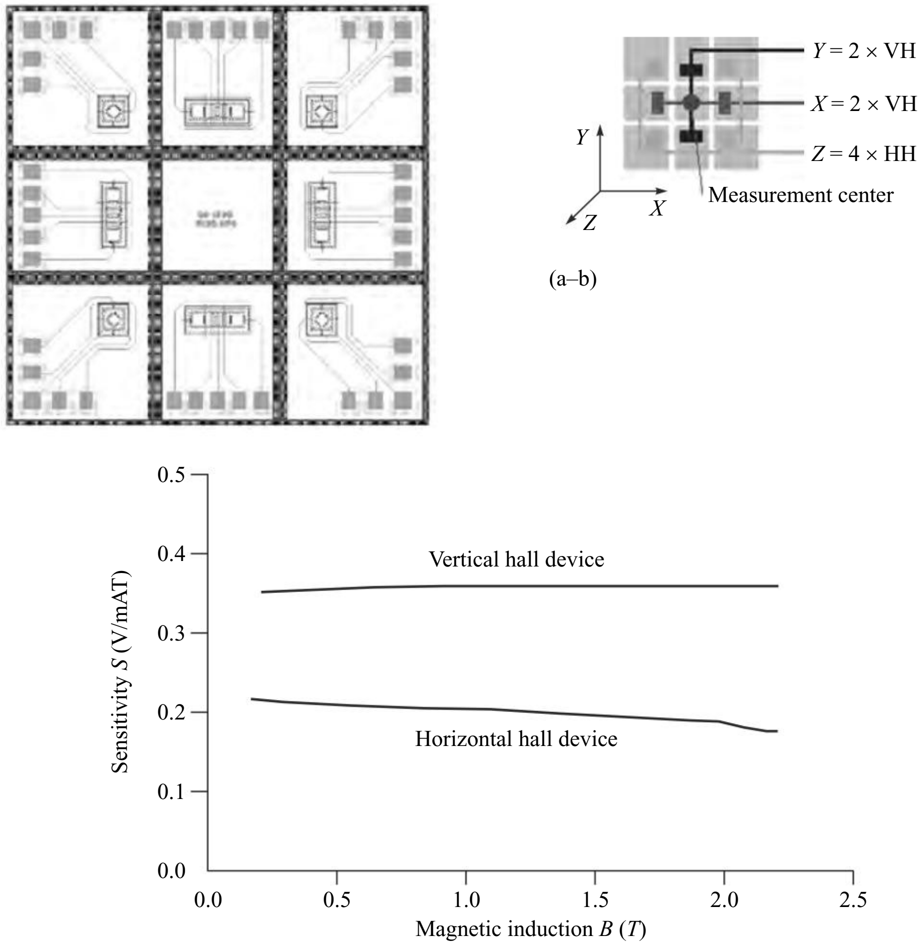
Magnetotransistor exhibits a linear magnetic response and can detect slowly or non-alternating signals. They are sensitive to magnetic inductions parallel to the chip surface, i.e. in-plane. This is complimentary to the perpendicularly out-of-plane sensitive lateral devices. The two collector terminals are designated as parallel and vertical terminal respectively, in order to capture in-plane and out-of-plane magnetic field simultaneously, if it is necessary. That is the vertical part of the current is used to measure the in-plane magnetic-field vector, while the lateral part of the collector current can be used to sense the component of the magnetic field perpendicular to the surface of the device. With additional processing circuitry it can also measure the relative magnitude of the magnetic field of both the planes.

In the general structure of the magnetotransistor the induced Hall field mainly results through diffusion carrier transport mechanism. The effect of carrier transportation by means of diffusion is quite small. The drift carrier transport does not play significant role. In order to increase the drift carrier transport novel magnetotransistor structure are under development. The novel structure can enhance the induced Hall effect and hence the sensitivity as well as the range of operation. This would facilitate lateral drift of carriers so that the three components of the field vector could be measured. Sometimes, multiple magnetotransistors each dedicated to measure respective component of the field vector are integrated in one wafer.

8.7.1 Integrated Magnetotransistor

There has been a pressing need to measure all the three components of a magnetic flux density vector at the same spot. Scott at Microsystems Institute developed a high accuracy microsystem for measurement of all three magnetic field components. The integrated system can also detect the direction of the field vector.

The microsystem consists of a total of eight Hall sensitive devices, named as subsensors, integrated on the same chip and divided into three groups. Each group measures one direction of the flux density vector. The magnetic flux density vector components x and y parallel to the sensor surface are measured each by one pair of the vertical Hall devices located on opposite sides of the microsystem, whereas the z component is measured by four horizontal devices located in the four corners of the system as illustrated in Fig. 8.9(a–b). The subsensors of each group are connected in parallel in order to produce a combined output signal that represents a linear interpolation to the center of the sensor. Figure 8.9(c)



(c)

Fig. 8.9 *(a–b) Structure and logical interconnection of the 3D microsystem for magnetic field measurement, (c) Sensitivity vs. magnetic induction of the vertical and horizontal devices in the 3D Microsystem*
Source: Scott, MIT

shows sensitivities of vertical and horizontal Hall devices. One can note that under same biasing conditions the vertical Hall device has sensitivity higher than the horizontal device. This requirement is essentially needed in such applications where one component of the magnetic field is much stronger than that of the two others. The important design criteria considered while fabricating such integrated devices are:

- No temperature-dependent mechanical stress should exert on the active regions of the subsensors when in operation.
- Good galvanic isolation of the subsensors so that the sensor can be used for applications in higher magnetic fields

- Cross-sensitivity needs to be eliminated

The advantages of integrated magnetic sensors are given below.

- The integrated technique can reduce the effect of noise and simplifies the signal-processing interface.
- The integrated system contains many electronic circuits such as amplifiers, biasing circuits, etc. for signal processing and calibration. The three signals can be read out separately.
- Integration of logic capability optimizes system performance.
- Detect magnetic field changes as small as 50 picoteslas; a million times weaker than the Earth's magnetic field.
- High spatial resolution
- Arrays of micron-sized magnetic subsensors can be fabricated for document validation applications including currency and credit cards and magnetic imaging
- Can be used in cryogenic environment
- A low-power consumption device
- Low-cost mass production of sensors

8.8 MEMS MAGNETIC SENSOR

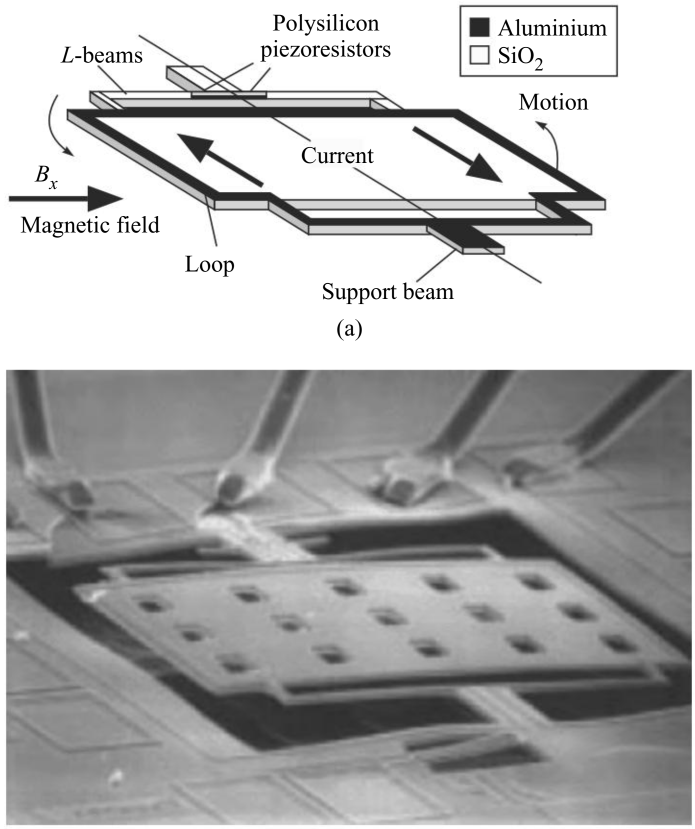
Semiconductor magnetic sensors based on silicon, described above, have inherent limitations to their sensitivity and resolution, restricting the performance. Using xenon diuoride etching of standard CMOS a novel magnetic field sensor can be fabricated based on a micro electromechanical systems (MEMS) approach. In this case the magnetic field vector is detected by measuring the amplitude of a mechanical Lorentz force. The MEMS device consists of a current loop on an SiO_2 plate. Amplitude is detected with a polysilicon piezoresistor Wheatstone bridge.

8.8.1 Construction

The MagMEMS devices are fabricated using the methods described in Chapter 2. The sensor consists of an oxide plate (SiO_2). The oxide plate is held suspended over an etched cavity in the silicon substrate by two sets of support beams. Etching is performed using xenon diuoride that frees the mechanical plate and enables it to rotate in real environment. Figure 8.10(b) illustrates the schematic diagram and the SEM photograph. The two support beams extend from the edges. Towards the end of the beams, a pair of *L* shaped beams are present at each side. The L-shaped beams are perpendicular to the support beams. Further, they travel the length of the plate and connect at the two ends on each side. The *L* beam on one side of the plate contains two polysilicon piezoresistors. The piezoresistors act as the active pair of a set of four resistors forming a Wheatstone bridge. The bridge transduces a change in resistance due to the strain experienced by the piezoresistors when the *L* beams bend due to a change of voltage. The voltage signal is then sent to electronics for processing. The set of beams on the other side of the oxide plate are primarily used to constitute a loop so that a conducting path is formed. The conducting path encircles the plate around its perimeter.

8.8.2 Principle

The direction of magnetic field, which needs to be measured, is shown in the figure. At the same time, an excited sinusoidal signal is applied through the current loop. The excited signal interacts with the



(a)

(b)

Fig. 8.10 *(a) Schematic diagram of Micromechanical Resonant Magnetic Sensor, (b) Its MEMS photograph*
Source: Beverley et al., *IEEE T on Device Let.*, 19/12, 1998.

magnetic field vector. This in turn produces a force, which acts on the plate. While the sinusoidal current is excited within the loop, if the device is kept in a transverse magnetic field, a force which tends to push up on one side of the plate and down on the other will be exerted. The mathematical expression of this force is given by:

$$F_s = I_L L_s B_x \tag{8.8}$$

where F_s is the force on one side of the plate, I_L is the loop current, L_s is the length of the side perpendicular to the magnetic field and B_x is the magnetic field along the plate. The oxide plate will thus rotate about the axis that is running through the support beams as shown, and develop a strain in the piezoresistors connected to the beam. If L_{ma} becomes the length from the end of the plate to the base of the piezoresistors, then the moment experienced at the base of the bending beams is given by,

$$M = 2F_s L_{ma} \tag{8.9}$$

Note that the moment is proportional to the area of the plate, A_p (in this case $A_p = 2\ L_s \times L_{ma}$). The strain experienced by the piezoresistors is:

$$\varepsilon = \frac{Mz}{EI_b} \tag{8.10}$$

Here, z is the distance from the neutral axis of the bending beam to the polysilicon piezoresistors, EI_b is called the flexural rigidity of the beams.

The change in resistance can be expressed as,

$$\frac{\Delta R}{R} = G\varepsilon \tag{8.11}$$

where, ΔR is the change in resistance in the strained piezoresistor, R is the original resistance, G is constant called gauge factor, which is approximately –20 in case of n-type polysilicon. The change in resistance appears as a change in voltage within the pre-balanced Wheatstone bridge. The voltage can be calibrated in order to reflect the value of magnetic field. The relationship between the change in resistance and change in voltage can be expressed as,

$$\frac{\Delta V}{V} = \frac{\Delta R}{2R} \tag{8.12}$$

8.8.3 Sensitivity

Sensitivity of a device plays a vital role in determining its capability and hence performance. This is a measure of ratio of change in transducer output to a change in the value of the measurand. For this particular sensor the sensitivity expression is given by,

$$S = \frac{\Delta V}{\Delta B} = \frac{GzI_L VA_p}{EI_b} \tag{8.13}$$

where ΔB is the change in magnetic flux vector. As can be seen, sensitivity depends on distance from the neutral axis of the bending beam to the polysilicon piezoresistors, gauge factor, loop current, area of the plate and flexural rigidity of the beams. Figure 8.11 shows the sensitivity curve of typical MagMEMS sensors.

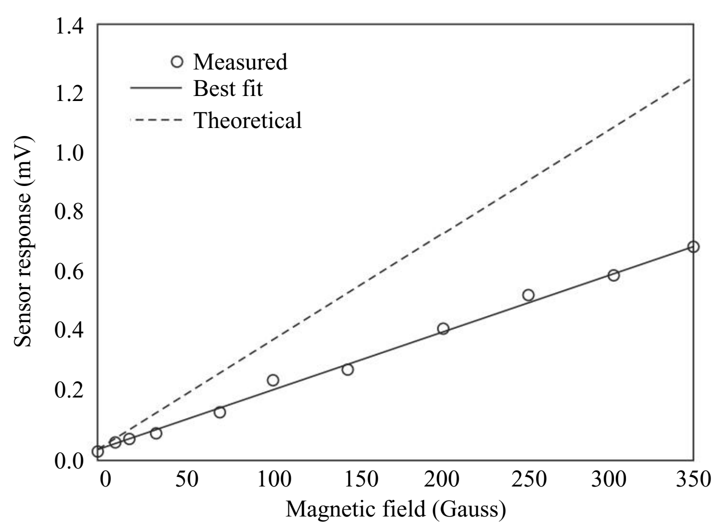


Fig. 8.11 *Sensitivity of the MEMS magnetic sensor*
Source: Beverley et al., *Copyright: IEEE T on Device Let.*, 19/12, 1998

8.9 PRESSURE SENSOR UTILIZING MOKE

The magnetoelastic material along with membrane technology is exploited to show proof-of-principle (PoP) based pressure sensor. The pressure sensor utilizes the principle of MOKE (**M**agneto-**O**ptic **K**err **E**ffect). The GaAs wafer is coated with SiN to act as a membrane (Fig. 8.12(a)). Using standard photolithographic technique and compatible etching method a rectangular membrane is opened out. The SiN is then coated with amorphous magnetoelastic material using the sputter deposition method.

The principle of operation is that when pressure is applied the permeability of the magnetoelastic materials changes. As the pressure deflects the membrane, its permeability is found to decrease (Fig. 8.12(b)). The permeability can be monitored utilizing the Magneto-Optic Kerr Effect. MOKE PoP corresponds to a change in the intensity or polarization state of light reflected from a magnetic material. Linearly polarized light experiences a rotation of the polarization plane. Thus, when a beam of polarised light reflects off a magnetised surface, the plane of polarisation of the light slightly rotates. The angle of rotation is called Kerr rotation. The PoP is sometimes referred to as surface magneto-optic Kerr effect (SMOKE), although usually the light penetrates about 20 nm into the surface for most material.

Based on the geometrical interaction of the surface and the incident light, there are three principal modes of operation for MOKE: Longitudinal MOKE (LM), Transverse MOKE (TM) and Polar MOKE (PM). LM geometry provides a signal proportional to the component of magnetization that is parallel to the film plane and the plane of incidence of the light. TM geometry provides a signal proportional to the component of magnetization that is parallel to the film plane but perpendicular to the plane of incidence of the light. Lastly, the PM geometry provides a signal proportional to the component of magnetization that is perpendicular to the film plane.

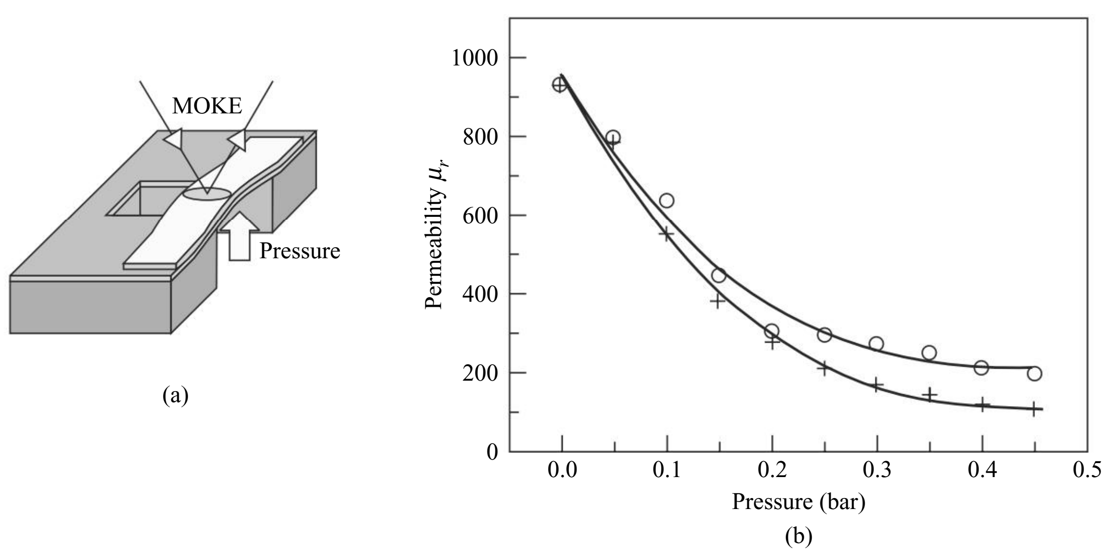


Fig. 8.12 *(a) MEMS magnetic pressure sensor, (b) Permeability versus pressure curve*
Source: Gibbs et al. *J of Physics (D)*, 37, 2004

8.10 MAGMEMS ACTUATORS

Many MEMS actuators are characterized by the following desirable parameters such as displacement resolution, long-range force, large-actuation output, low-power consumption, low-voltage requirement, environment compatibility and linearity.

MagMEMS actuating devices have several advantages over their counterpart electrostatic types in terms of generation of long-range force and deflection with low driving voltage in thoughtless environment. Actuators with in-built permanent magnet have benefits from low power consumption, good resolution and simple integration. Permanent magnet can provide a desirable constant magnetic field without the consumption of electrical energy. Other important characteristic of the MagMEMS actuator is that it does not generate heat. Further, the energy stored in a permanent magnet does not deteriorate if the device is micromachined appropriately satisfying the precision design criteria and handled properly during its use. The non-deterioration of energy is due to the reason that it does no net work on its surroundings. Moreover, MagMEMS actuators can achieve relatively high energy density in microstructures, compared to other energy storage devices in the microscale.

MEMS microactuators utilizing magnetic proof-of-principle are designed by using bulk or surface micromachining technique. The actuators are usually based on out-of-plane rectangular flap configuration with cantilever beams that are deflected by magnetic force generated from either an electromagnetic coil or microstructure deposited with Permalloy. Permalloys ($Ni_{80}Fe_{20}$) are nickel alloys containing about 20 to 60% Fe, used for achieving high magnetic permeability and electrical resistivity. The low-stress electroplated Permalloy material acts as the medium for magnetic interaction and force generation. Large deflections of typically greater than 100 micrometer in the presence of either an externally applied or a locally induced magnetic field can be achieved.

8.10.1 Optical Switches

A switch is a device consisting of a mechanical, electrical, electronic or optical element for making or breaking or changing the connections in a circuit. In optical communication a switch routes the optical signal without electro-optical and opto-electrical conversions. These critical components are used in optical networks for interconnecting of one or more devices in the network. Conventional mechanical switches have been used for over 20 years. They are large in size, slow in switching speed and unreliable for high performance applications. Solid-state switches are small but have inherent disadvantages of high loss and high cross talk. The rapid growth of high-speed optical communication networks calls for development of low power, small volume, and low cost optical switches. MagMEM optical switches have advantages over solid-state switches as they offer higher isolation, low insertion loss, low power consumption and improved linearity. They are also faster in comparison to traditional electromechanical switches and relays.

The magnetic actuation for switch application is called *latching*. Latching eliminates the need for the static power supply and that can latch ON or OFF with zero power required to hold this state. The switch can be nonvolatile and bistable. Magnetic switch, primarily consists of four components as given below.

- Cantilever
- Permanent magnet
- One-dimensional (1-D) planar coil
- Electrical contacts

The cantilever is supported by torsion flexures. It is essentially a two-layered composite. It consists of magnetic material like Permalloy on the top and a highly conductive material like gold at the bottom

surface. The switching contact end, called free-end appears on the right of the cantilever. Free-end can be made to deflect up or down by applying appropriate current through the coil. The switch is OFF when no current flows. When deflected, the cantilever makes electrical contact with the bottom conductors and the switch is said to be ON, i.e. the loop is closed. The permanent magnet holds the cantilever up or down after switching, making the device a latching relay. Single-pole-double-throw and RF magnetic MEMS switches are usually designed by this principle.

8.10.2 An Example of 2 × 2 Optical Switch

A schematic diagram showing the principle of operation of a typical 2 × 2 micromachined structure of a magnetic actuation-based optical bypass switch is illustrated in the Fig. 8.13. A micromirror makes it possible to connect and disconnect the optical path. It is a double-sided thin micromachined MEMS mirror. When the mirror is brought up (ON) as shown in Fig. 8.13(a), both the input optical signal will be reflected and will pass through the neighboring output optical fiber as shown. When the mirror is brought down (OFF), as shown in the Fig. 8.13(b), the two input optical signals would pass in the direction as shown. The movement of the mirror is performed by electromagnetic actuation. Typical dimension of the mirror is about 2 mm long, 300 μm high and 5 μm thick. The displacement of mirror is about 200 μm.

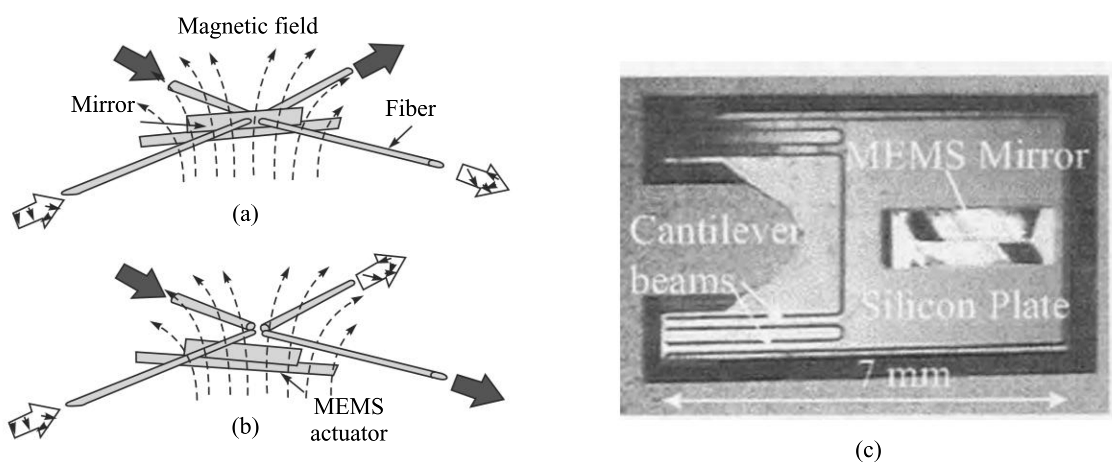


Fig. 8.13 *Principle of switching (a) Switch ON, (b) Switch OFF, (c) A micromachined optical switch*
Source: Miller et al., California Inst. of Tech.

The bottom view and the dimensions of the mirror are shown in Fig. 8.14. The mechanical structure of the switch is that it has a silicon plate. Four 35 μm wide and 20 μm thick cantilever beams support the plate. Bonding pads are essentially required at the base to hold the beams with the substrate. The electromagnetic actuating component is a planner coil. The copper coil is electroplated and placed on the bottom side of the silicon plate. The entire integration is then placed on the rare-earth magnet as shown in Fig. 8.15. The rare-earth magnet provides a constant but large magnetic field gradient. The actuation is based on the driving current in the coil. When current is passed through the coil, it interacts with the magnetic field resulting a force that exerts upon the plate causing it to deflect along with the mirror. A current of 30 mA, typically, can lift the mirror up to the desirable typical deflection of about 200 μm. The mirror can be fabricated by TMAH (Tetramethylammonium Hydroxide) etching a (110) silicon

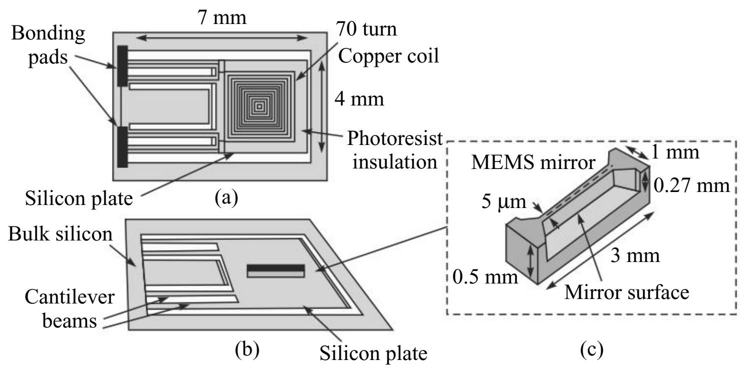


Fig. 8.14 *(a) Bottom view of the actuator, (b) Side view, (c) MEMS mirror on silicon plate*
Source: Miller et al., California Inst. of Tech.

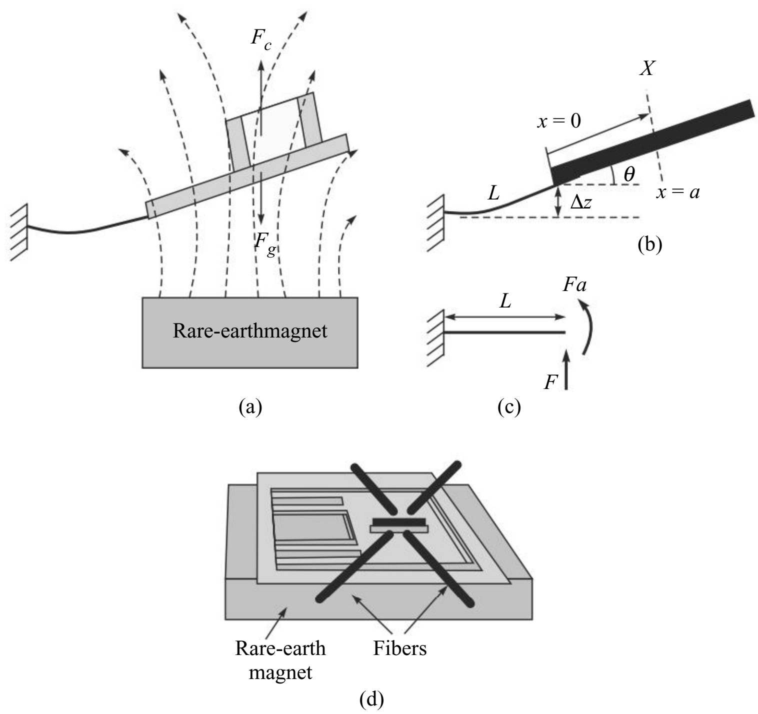


Fig. 8.15 *(a) Force acting on cantilever, (b) Deflection, (c) Mechanical representation, (d) Schematic diagram of assembled 2 × 2 optical bypass switch*
Source: Miller et al., California Inst. of Tech.

wafer with a patterned silicon-nitrite mask. The desired performance parameters while fabricating the switch is that the mirror surface should be smooth and thin. The smoother surface will avoid signal loss. The assembly procedure is as follows.

- Integration of the mirror on the top side of the plate
- Placement of coils above the rare-earth magnet
- Alignment and fixing of four optical fiber

For modeling the optical switch let us define the following terms.

F_c = Electromagnetic force acting on the cantilever
I_c = Coil current
m = Effective magnetic moment produced the coil
A = Total area of the coil
dB/dz = Magnetic field gradient in the z-direction
F_g = Gravitational force on the actuating element
g = Gravitational acceleration
ρ_{Si} = Density of the silicon
ρ_{Cu} = Density of the copper
V_{Si} = Volume of the silicon plate
V_{Cu} = Volume of the copper coil
Δz = The deflection at the tip of the beam
L = Length of the beam
a = Distance from the tip of the beam to the center of the silicon plate
E = Young's modulus for silicon
J = Moment of inertia of the beams combined together.
θ = Angle of deflection

Now the electromagnetic force acting on the cantilever, the gravitational force on the actuating element, the deflection at the tip of the beam and the angle of deflection are given by the following expressions, respectively.

$$F_c = m\frac{dB}{dz} = AI_c\frac{dB}{dz} \tag{8.14}$$

$$F_g = g(\rho_{Si}V_{Si} + \rho_{Cu}V_{Cu}) \tag{8.15}$$

$$\Delta z = (F_c - F_g)\left(\frac{L^3}{3\,EJ} + \frac{aL^2}{2\,EJ}\right) \tag{8.16}$$

$$\theta = (F_c - F_g)\left(\frac{L^2}{2\,EJ} + \frac{aL}{EJ}\right) \tag{8.17}$$

8.11 BIDIRECTIONAL MICROACTUATOR

Microactuators can be driven in both directions. Then these actuators are called bidirectional actuators. There has been a growing interest in the realization of bidirectionally driven microactuators in the areas of optical communication and computing devices, such as scanners, etc. The bidirectional actuation can also be realized by adopting electrostatic and thermal actuation methods. This section deals with the principle of magnetically driven bidirectional actuation. Magnetically driven actuation mechanism is appropriate for a large deflection over several hundreds of micrometers. Bidirectional actuation is achieved between the permanent micromagnet and the electromagnet.

In place of permanent magnet variable reluctance-type electromagnetic can also be used, however, the latter type does not have significant scaling factor (resolution) and also consume more power. The exciting current through the electromagnet makes it possible to attract the permanent micromagnet

causing deflection and hence actuation. When the direction of the exciting current is altered the actuation occurs in the reverse direction. This means that by altering the direction of the exciting current through the electromagnet, either attractive or repulsive force can be generated between the two types of magnets. The bi-directional actuator is also known as configurable actuator.

A schematic diagram of such microactuator is illustrated in Fig. 8.16. Besides the permanent magnet and electromagnet, the other important element of actuator is the silicon cantilever beam. The tip of the cantilever actually accommodates the array of permanent magnets. The array of permanent magnets is electroplated in order to improve and control the vertical anisotropy. Electroplating process is considered essential in constructing the permanent magnet arrays. Ceramic magnets and a paddle agitator are used to build the electroplating system. Ceramic and paddle agitator can also improve the controllability as far as vertical anisotropy is concerned. Further, this reduces the irregularity of electroplated materials. The shape of the electroplated permanent magnet is either cuboids or rectangular. Typical dimension of the silicon cantilever beam with permanent magnet arrays has length of 5 mm, width of 1 mm and thickness of 10 micrometer.

In order to prepare the cantilever, a silicon wafer is chemically etched in agitated Tetramethylammonium hydroxide solution. Time controlled etch stop is used to decide the etching depth. Then, the magnet arrays are fabricated over the silicon island of dimension 2 by 2 mm. Finally the silicon cantilever beam is released from the other side by using the reactive ion etching (RIE) technique.

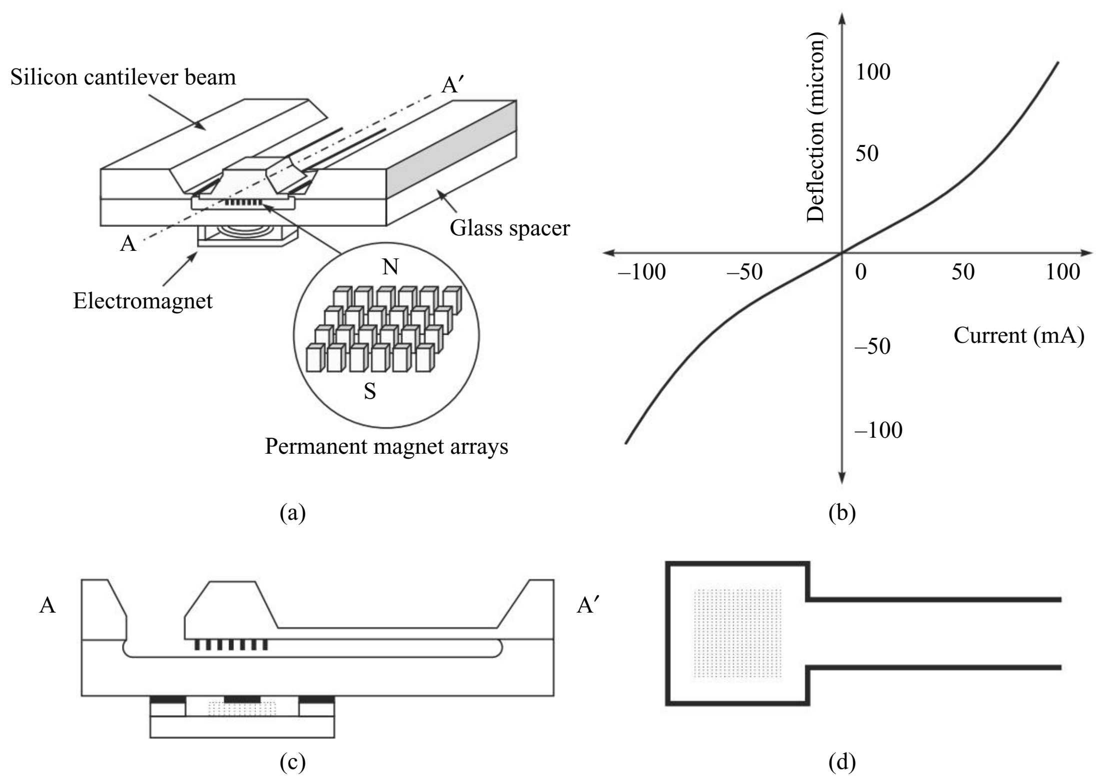


Fig. 8.16 *(a) Cantilever structured magnetic bidirectional microactuator, (b) Bidirectional deflection of a microactuator as a function of current in electromagnet, (c) Another view, (d) Magnet arrays: Source:* Cho, *J of MEMS*, 11/1, 2002

The design steps consist of only two phases: electroplating permanent magnet arrays and releasing of the cantilever beam from the silicon substrate. A detail illustration of the fabrication process is presented in Fig. 8.17. The important steps are,

(a) Oxidation
(b) Silicon etching
(c) Re-oxidation
(d) Electroplating permanent magnet arrays
(e) Releasing cantilever beam by RIE and stripping metals and oxide
(f) Cr and Au deposition
(g) Patterning and hard baking of photoresist
(h) Etching Au and Cr
(i) Etching glass in HF
(j) Removing metal layers
(k) Assembling a cantilever beam, a glass spacer and an electromagnet.

The electromagnetic force between a permanent magnet and the electromagnet is given by,

$$F = V \cdot M_z \cdot \frac{\partial B_z}{\partial z} \tag{8.18}$$

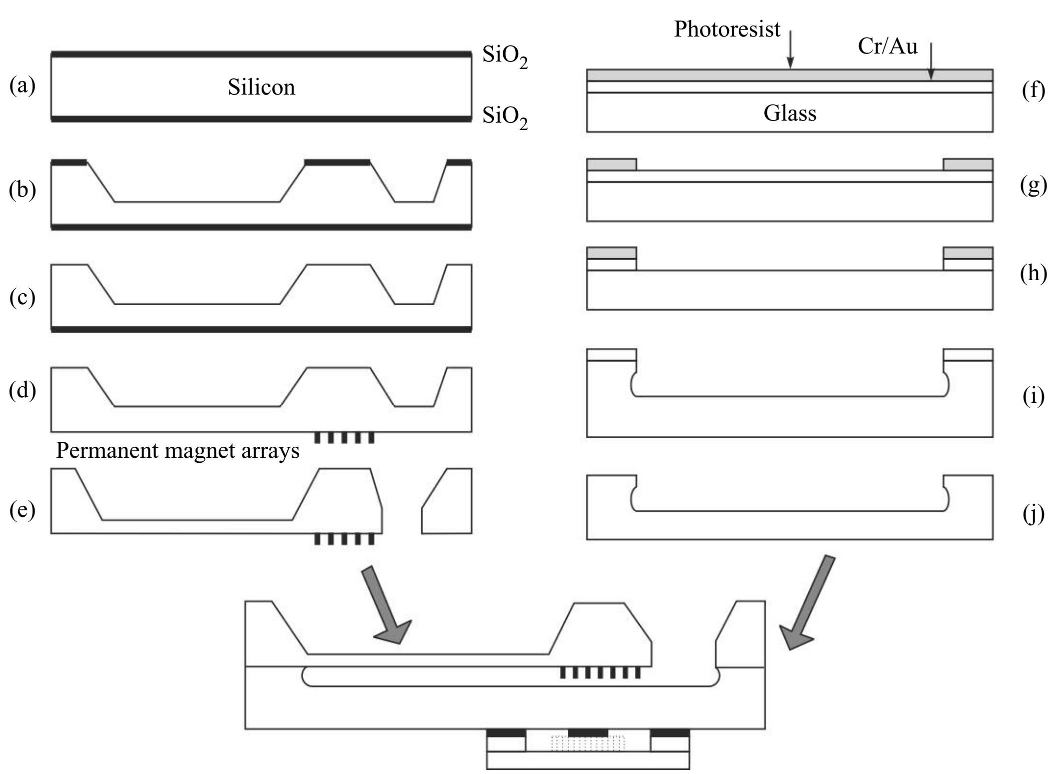


Fig. 8.17 *Fabrication steps of a bi-directional cantilever beam magnetic actuator*

where V is the volume, M_z is the magnetization of the magnet, and B_z is the flux density of the electromagnet. The static deflection of a cantilever beam, can be expressed,

$$\delta = \frac{2Fa^2}{Ewt^3}(3L - a) \tag{8.19}$$

where L is the length, E is Young's modulus, w is the width and t is the thickness of the silicon beam; a is the distance from a fixed end of the cantilever beam to a point of maximum magnetic force, where the center of magnets are located.

Figure 8.18 shows another structural example with two cantilevers, in which electroplated Permalloy is also used. Let us define the terms as follows;

L = Length of the magnetic piece
W = Width of the magnetic piece
T = Thickness
l = Length of the cantilever beam
w = Width of the cantilever beam
$a = w/2$
t = Thickness of the cantilever beam
$b = t/2$
H_{ext} = External magnetic field (normal to the plane of the plate structure)
H_1 and H_2 = Magnetic field strengths at the top and bottom edges of the plate
M = Developed magnetization vector in the permalloy piece
M_s = Saturation magnetization
F_1 and F_2 = Generated two force components due to the external magnetic field
$F = F_2 - F_1$ = Development of force because of interaction of H_{ext} and M
M_{mag} = Corresponding development of torque because of interaction of H_{ext} and M
H_k = Value of external magnetic field at which the saturation of magnetization just occurs
x = Horizontal coordinates of the point along the cantilever
y = Vertical coordinates of the point along the cantilever
E = Young's modulus of the cantilever
I = Moment of inertia of the cantilever
s = Arc length of the curvature (Fig. 8.18(f))
r = Radius of curvature
θ = Angular displacement experienced by each torsion beam
G = Torsion modulus of elasticity of the material
k = A constant determined from the cross-sectional geometry of the beam
θ_f = Maximum angular displacement
y_f = Maximum vertical deflection

Now the governing model equations of the microactuators can immediately be written. The magnitude of the two force components, one acting at the upper edge and other acting at the lower edge are given by,

$$F_1 = M_s \cdot W \cdot T \cdot H_1 \tag{8.20(a)–(b)}$$
$$F_2 = M_s \cdot W \cdot T \cdot H_2$$

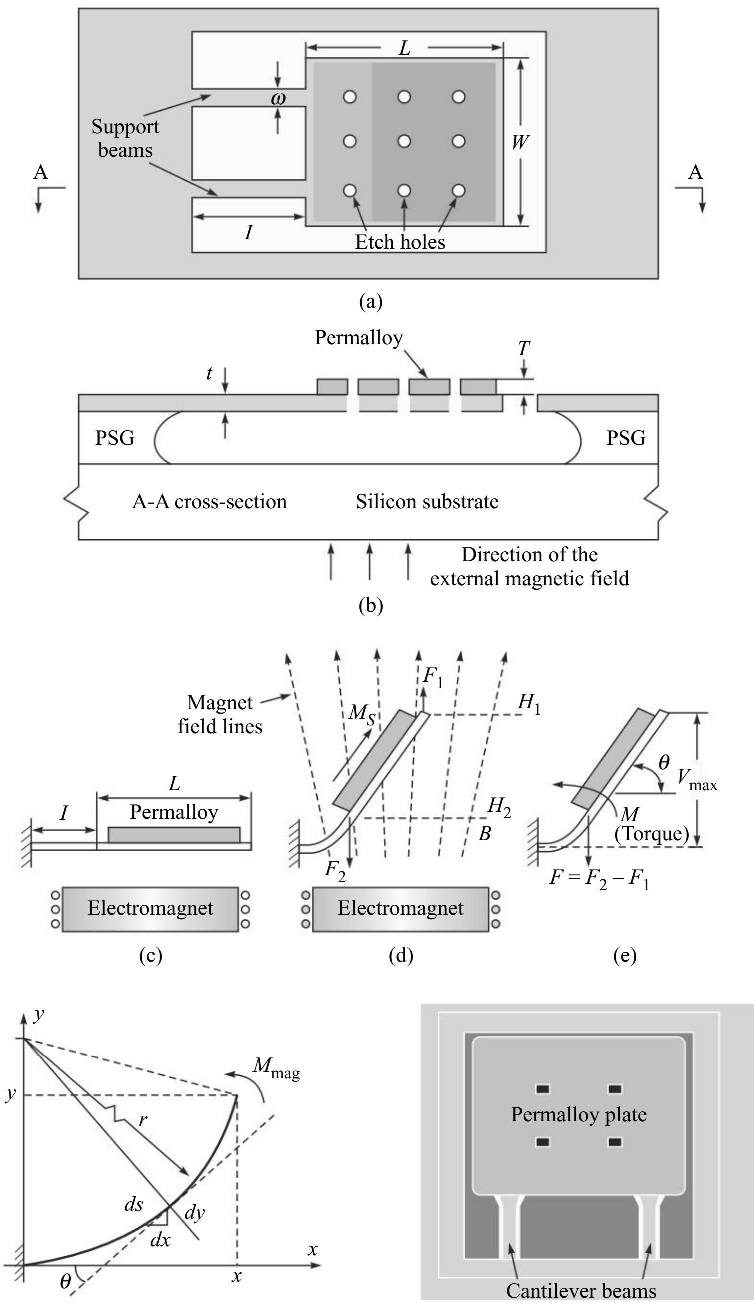


Fig. 8.18 *(a) Magnetic actuator with two cantilever-beam supports: top view and (b) Its side view, (c) Rest position when $H_{ext} = 0$, (d) Out-of-plane actuation, (e) Force system acting on the free ends of the cantilever beams, (f) Flexure cantilever beam under the magnetic torque, (g) Micromachined magnetic actuator*

Source: Liu, *IEEE T on Magnetics*, 35/3, 1999

Note that $F = F_2 - F_1$. Now, the developed counterclockwise torque because of interaction of H_{ext} and M, which acts on the bottom edge of the structural plate is expressed as,

$M_{mag} = F_1 L \cos\theta$, which is again equal to 2 $kG\theta/l$, where, k is given by,

$$k = ab^3\left[5.33 - 3.36\,\frac{a}{b}\left(1 - \frac{b^4}{12a^4}\right)\right] \tag{8.21}$$

The maximum angular displacement and maximum vertical deflection, both occurring at the free end of cantilever beams, are expressed as,

$$\theta_f = \frac{(\pi/2 - 1)FR^2}{EI} \qquad (8.22(a)–(b))$$

$$y_f = \frac{(3\pi/4 - 2)FR^2}{EI}$$

8.12 FEEDBACK CIRCUIT INTEGRATED MAGNETIC ACTUATOR

Surgical equipment and system developers have focused on minimizing their equipment in order to achieve precision surgical procedure in terms of minimally invasive surgery (MIS). MIS takes the advantage over the traditional open surgery with less pain, medical cost and rehabilitation time. Through advances in technology, MagMEMS actuation systems are being developed for surgical applications. Son and Lal at University of Wisconsin developed a magnetic vibrating microactuator that can be used for microsurgery applications (see Fig. 8.19). The actuator is remotely controlled by employing piezoresistive feedback system.

The actuator consists of an electromagnet, ferromagnetic mass and silicon nitride cantilever. In conventional magnetic actuator fabrication, a high-aspect lithography combined with electroplating is required to make high aspect ratio structures. However, they fabricated high aspect ratio columns with a suspension of ferromagnetic nanoparticles and epoxy using magnetic extrusion. The fabrication process did not require lithography.

The remote magnetic actuator with feedback consists of two basic units: actuation unit and feedback unit. Ferromagnetic mass is placed on the tip of the cantilever. The electromagnet is placed above the ferromagnetic mass as shown in Fig. 8.19. If AC current is applied to the electromagnet, the magnet repeats pulling and releasing the cantilever. If the cantilever is released, the bent cantilever returns to the initial position by its own spring force. One implied factor is that if the frequency of current applied to the electromagnet equals the resonance frequency of the cantilever the actuation is amplified by the Q factor of the mechanical resonance. The resonant frequency of the cantilever with mass can be varied due to the change of its mechanical boundary conditions. The feedback unit consists of three sub-units namely, strain gauge, amplifier and VCO (Voltage Controlled Oscillator). When the cantilever is vibrated at different frequency, the value at strain gauge decreases. The drop in the value of strain gauge triggers the voltage-controlled oscillator to adjust the frequency using linear feedback to find the new resonance frequency.

When the electromagnet is placed near the ferromagnetic column, the cantilever feels significant bending, causing displacement. The displacement can be calculated from the amplified voltage output from the piezoresistors strain gauge fitted on the cantilever.

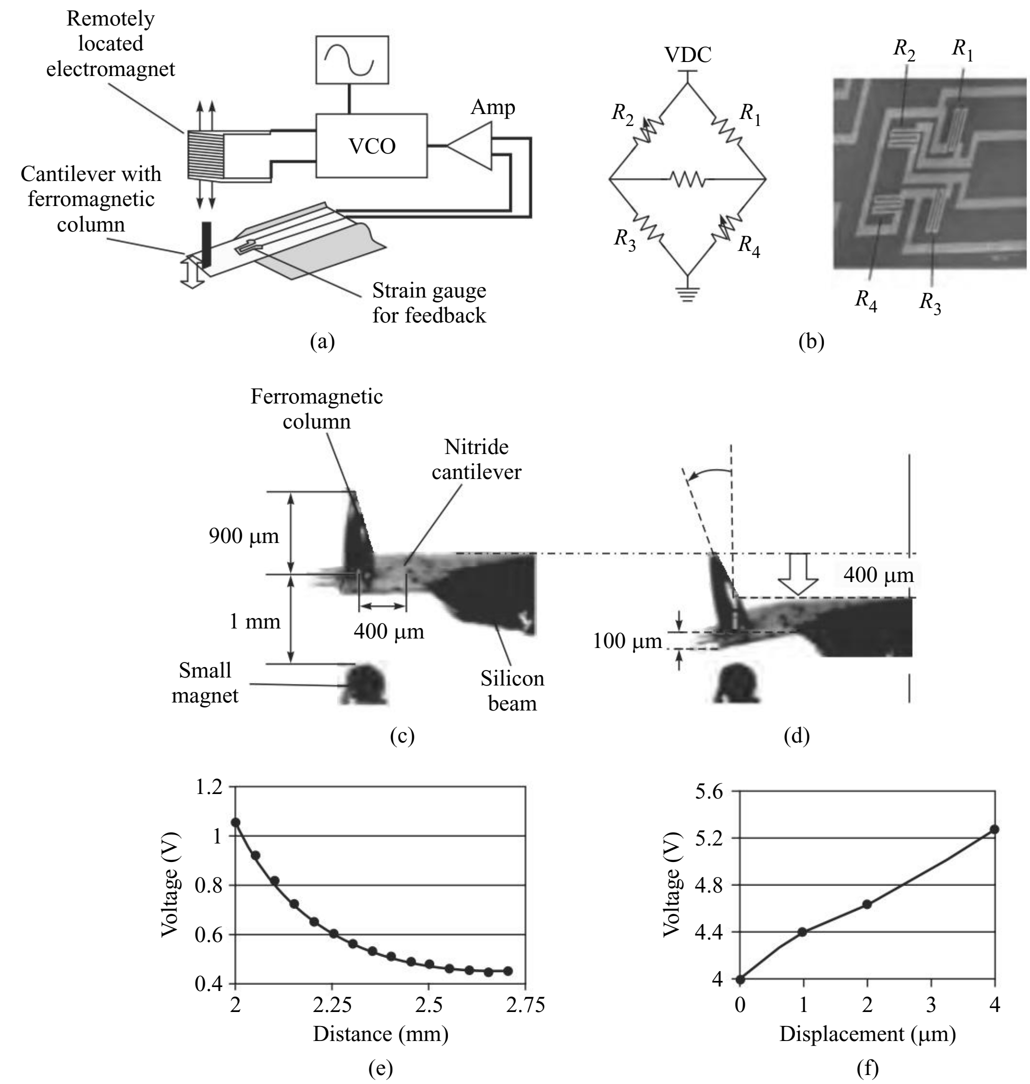
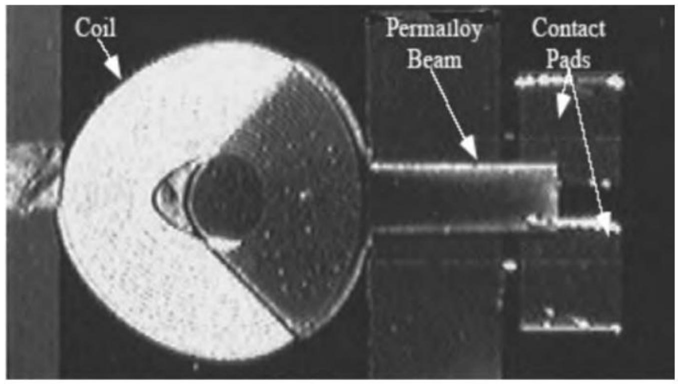


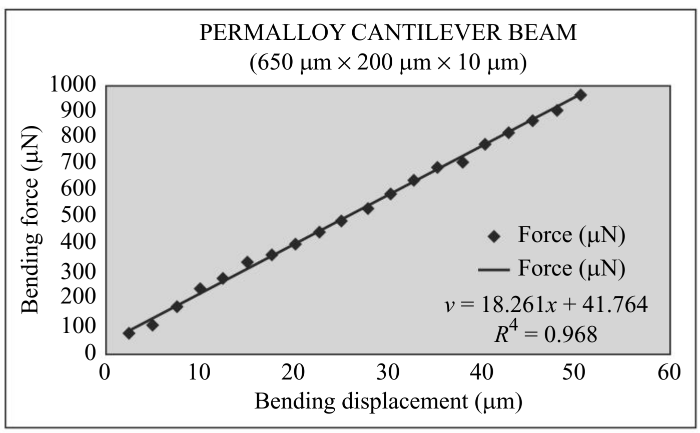
Fig. 8.19 *(a) Architecture of a typical remotely actuated magnetic vibration actuator, (b) Wheatstone bridge arrangement for integrated piezoresistors with MEMS picture, (c) Rest state, (d) Actuation (bending of the cantilever due to the magnetic force), (e) Distance versus voltage curve—When the distance between a permanent magnet and end-effecter is varied, (f) Amplified strain gauge output when the cantilever is displaced by known distance*
Source: Son and Lal, U of Wisconsin

8.13 LARGE FORCE RELUCTANCE ACTUATOR

A large-force, fully-integrated, electromagnetic actuator for microrelay applications is reported by Wright. The fabrication of the actuator is compatible with CMOS processing technology. The large force reluctance actuator is designed for high efficiency actuation. Figure 8.20 shows the top view of



(a)



(b)

Fig. 8.20 *(a) Top view of a reluctance actuator, (b) Displacement versus force curve (Write et al., Transducers'97, 1997).*

the microactuator. The actuator integrates a cantilever beam and planar electromagnetic coil into a low-reluctance magnetic circuit using a combined surface and bulk micromachining processes. The test result shows that a coil current of 80 mA generates a 200 μN actuation force. Theoretical extrapolation of the data indicates that an actuation force in the millinewton range can be produced by a coil current of 800 mA. The coil is integrated into the MEMS architecture capable of providing millinewton force with large displacement. In this actuator the concentration has been given on simultaneous optimization of magnetic permeability μ, air gap reluctance R_{gap}, magnetic core reluctance R_{core} and also beam spring constant k. The actuator is sometimes referred to as reluctance type actuator as, typically, the deflection force depends on the magnetic reluctance.

From Maxwell's equation, the force generated at the air gap is given by,

$$|F_{\text{mag}}| = \frac{dE_{\text{mag}}}{d} = \frac{1}{2}\left(\frac{NI}{\Re}\right)^2 \frac{d\Re}{d} \tag{8.23}$$

where, $\Re$ is the magnetic reluctance and is given by,

$$\Re = \frac{l_{\text{core}}}{\mu_{\text{core}} A_{\text{core}}} + \frac{l_{\text{gap}}}{\mu_{\text{gap}} A_{\text{gap}}} \tag{8.24}$$

By substituting the magnetic reluctance in previous equation we have the force expression, which can be written as,

$$|F_{mag}| = \frac{1}{2\mu_o A_{\text{gap}}} \left(\frac{NI}{\Re_{\text{beam}} + \Re_{\text{gap}}} \right)^2 \tag{8.25}$$

where,

F = gap magnetic force
$\Re$ = magnetic reluctance
N = number of coil-turns
I = coil current
A = Cross-sectional area
μ = magnetic permeability
l = length

From the above equation it can be noted that large forces require low magnetic circuit reluctance and large amp-turns. Minimizing magnetic reluctance entails use of a material having large permeability. A low μ material can cause the reluctance of the magnetic circuit to become dominated by the magnetic material and not the air gap. To maximize the efficiency of the actuator, it is desirable to have $\Re_{gap} >> \Re_{core}$.

8.14 MAGNETIC PROBE BASED STORAGE DEVICE

Magnetic probe based storage device utilizing MEMS technology is a promising technology, which can store more than 2 GB of data on 2 cm^2 of die area. The fabrication is simple and also compatible with a standard integrated circuit manufacturing process. The technology accentuates a thin film magnetic storage medium that is positioned in the *XY*-plane, as shown in Fig. 8.21. The writing is achieved magnetically by means of an array of clubbed probe tips. Each tip is actuated in the *Z*-direction. Unlike disc drive, the Media Sled (MS) moves along the *X* and *Y* axes to seek to new locations. There can be an array of several thousand of fixed probe tips, typically 10,000 probe tips under a media sled area of 1 cm^2. By using the combination of optical lithography and e-beam lithography to pattern a layer, 50 nm diameter probe of a permanent magnetic material with a height of 100 nm can be created. The mechanism for reading and writing data bits requires the probe tip to be extremely close to the MS. The probe tips are permanent magnets, which during operation access the rectangular storage MS by the process of scanning. Each probe tip accesses a specific rectangular region on the media and no two such regions overlap.

Data are written (stored) and retrieved by positioning and moving the MS over the probe tip array. The MS is a thin film magnetic media and the tips are permanent magnetic tips. The media resides on the top and the probe array on the bottom wafer. There are comb-like microactuators which drive the media sled in *X* and *Y* directions. There are independent actuators, which drive the media sled in *Z* direction. These are called head actuators. Each permanent magnetic probe recording head is usually carried by an independent parallel plate electrostatic actuator, which controls its head-to-media spacing. The MS itself acts as one of the electrode of the parallel plate electrostatic actuators. It is common to all the head actuators. To cover a large media area, the MEMS device usually needs a large number of probe tips and head actuators. The MS is movable and is suspended by springs. The anchors attach the springs to the wafer.

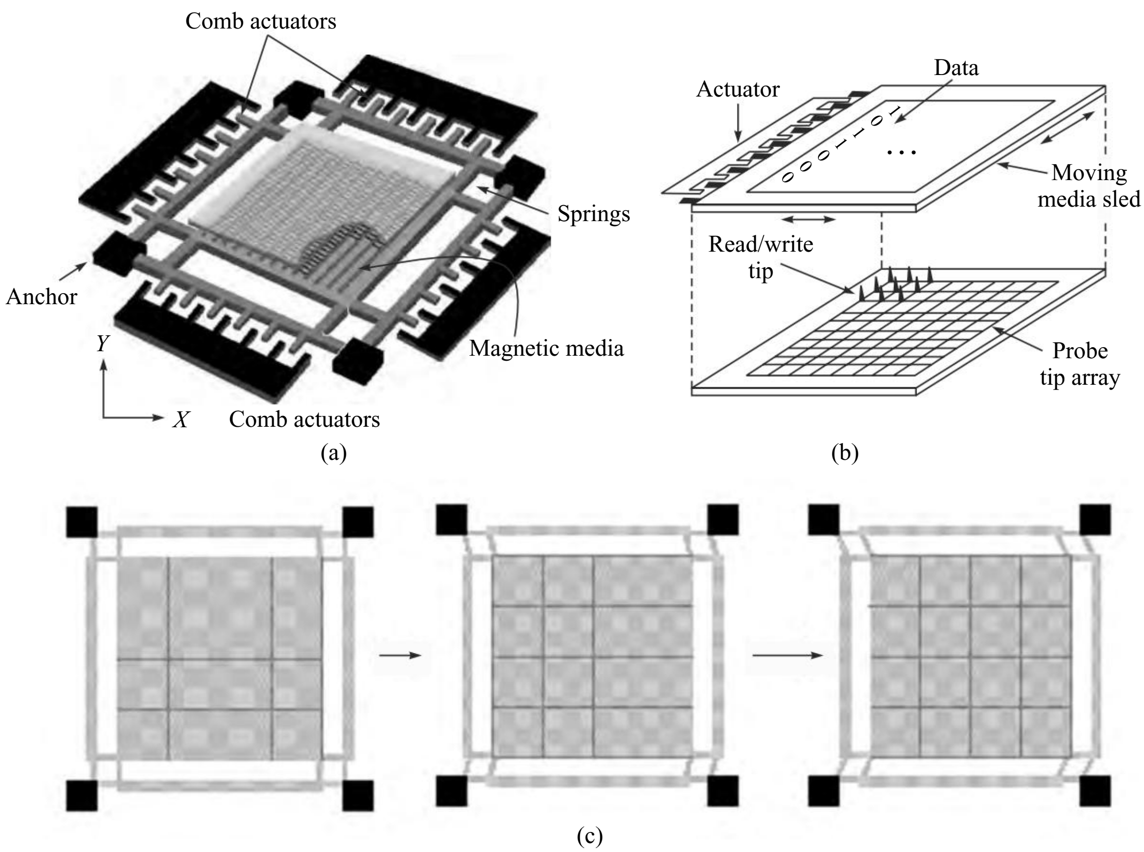


Fig. 8.21 *(a) Schematic diagram of magnetic storage device (The figure has been shown up-side down for clarity, i.e. the media is shown below and the probe tips are on the top contrast to that is shown in Fig. b); (b) Detail view of the probe tips (c) Examples of actuations movement of the sled in the X-axis. (Uysal et al., Proceedings of the 2nd USENIX Conference, 2003)*

Writing is carried out with the probe tip by simply modulating the spacing between the tip and the media. Reading of data is accomplished by knowing the magnetic state of the media, which is achieved by sensing the force that the media exerts on the permanent magnet head. Note that the interaction force between a permanent magnetic tip and a thin film magnetic MS is about hundreds of pN. A probe tip to media spacing typically of 2 nm can write (store) the bits reliably while a spacing of 11 nm can effectively read the data from the MS.

Figure 8.22(a) illustrates a very small portion of the magnetic media sled (Typical specification of magnetic storage sled is given in a table in Appendix B) As can be seen, the magnetic media on the sled is organized into rectangular regions, called *tip regions*. For illustration purpose only 16 (4 × 4) rectangular regions are shown, although typically thousands of tip regions are common. Each rectangular tip region area has N column and M rows and can be able to store N×M bits. Each rectangle outlines the area accessible to a single probe tip. The tip regions are linearly numbered such as A1, A2, A3, A4, and so on. A column in the rectangular region is called *tip track*. A tip track has M bits and is organized into several vertical *tip sectors*. Tip sectors contains servo information and encoded data.

Note that like standard CD operation, data is not byte-accessible and the smallest accessible unit of data area is the tip sector. A tip track is divided into multiple *logical sectors*. In other words a logical sector is formed when many tip sectors are grouped. In this typical example, as shown in Fig. 8.22(d), each logical sector consists of two tip sectors. In particular, the logical Sector$_1$ as shown is composed of the first tip sectors of the two upper tip regions such as A1 and A2. A *cylinder* is defined as all of the columns of data with the same X coordinate. It consists of all tip sectors that are accessible by any tip without moving the sled along the X-axis. Note that a track is a subset of a cylinder consisting of all tip sectors that can be accessed simultaneously by the active tips. To read or write data, the media passes over the active tip(s) in the $\pm Y$ direction while the tips access the media. A track consists of all bits

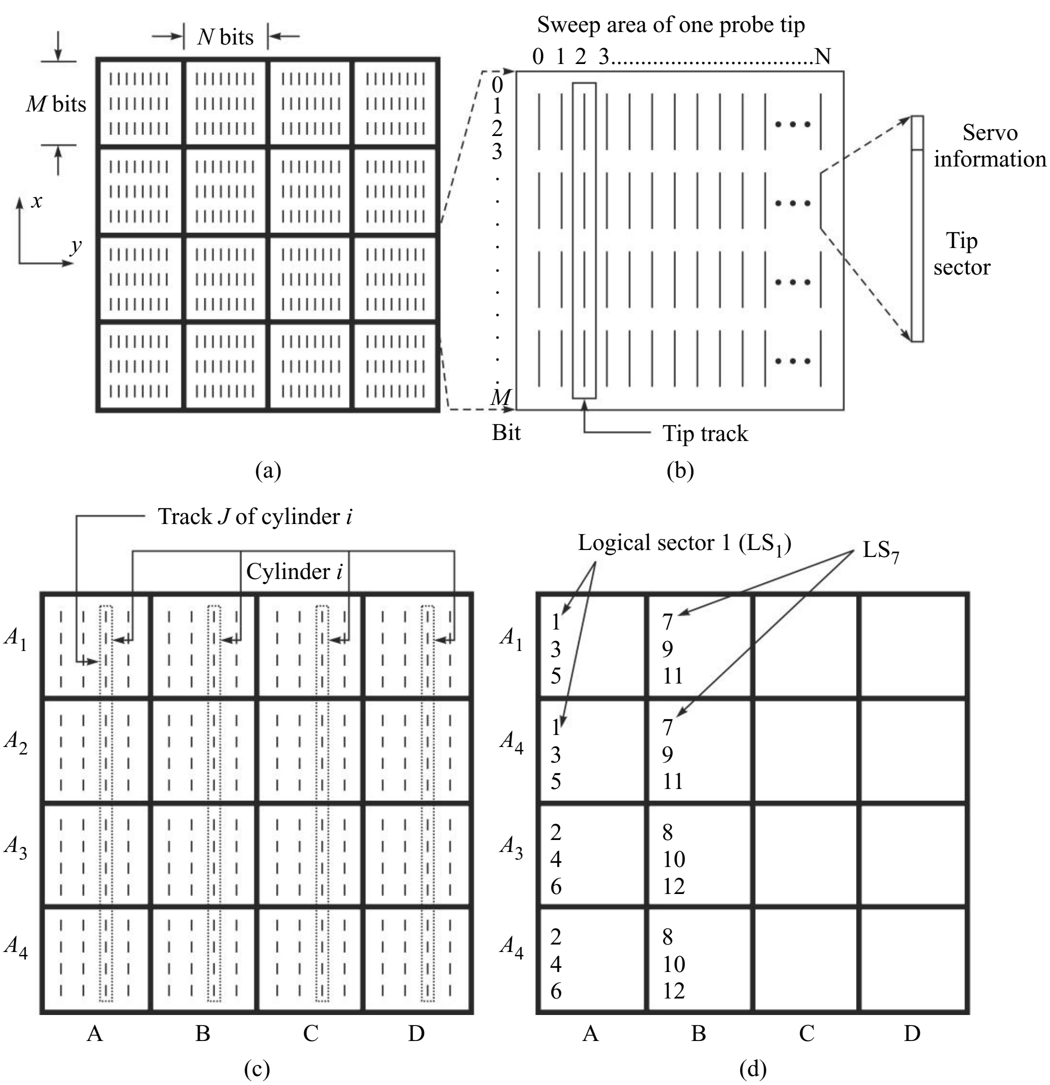


Fig. 8.22 *(a) Illustration that shows rectangular region, sweep area of one probe tip, tip track, tip sector (Source 55), (b) detail view, (c) and (d) Track, sector and cylinder (Griffin et al., Proc. of ACM SIGMETRICS, 2000; Courtesy:* ACM SIGMETRICS*)*

within a cylinder that can be read or written by concurrently active tips. Figure 8.22 shows the 4×4 rectangular regions, the sweep area of one probe tip, tip track and the tip sector.

Each bit can be represented by something called a *triple* such as <*x*, *y*, tip>. Here <*x*, *y*> represent the coordinates of the bit of interest and the <tip> signifies the probe tip within that addressable region. A tip tract can be read or written by the active tip as the magnetic media sled moves along the *Y* direction. A tip track contains M bits, each with identical values for <*x*, tip>. In Fig. 8.22, tips A1, A2, A3 and A4 are active and the corresponding track is indicated.

MEMS magnetic probe based storage device can access gigabytes of high-density, nonvolatile magnetic storage data and approximately more than 6 times faster than that of a modern disk. It is usually fabricated on the surface of silicon wafer. The actuation is achieved in response to electrostatic, thermal or electromagnetic forces. Because of their extremely small size, the MEMS based mass storage devices can be useful in applications such as appliances, sophisticated teaching toys, biomedical monitoring devices, civil infrastructure micro- and nano-satellites, etc. MEMS magnetic probe based storage device has several important consequences as given below.

- Offers data storage of very high densities compared to standard disk system
- It does not depend on continuous availability of a power source
- Stored data is persistent
- It only involves positioning delays. No rotational delay is encountered. The positioning delay depends on the relative positions of the sled and of the destination coordinates. This delay is much smaller than in disk drives
- Typical access times for MEMS are in the order of 1–2 ms
- Ranges of motion are in the order of a few millimeters
- Components have small masses

8.15 SUMMARY

Micromachining techniques are used to realize micromagnetic devices. Micromagnetic devices are of three types: sensors, actuators and date storage devices. Magnetic sensors are used for the detection of current and magnetic field. Magnetic materials are used to fabricate the sensor. There are three types of magnetic materials: magnetoresistive, magnetostrictive (also called magnetoelastic) material and permanent magnetic material. Magnetoresistive material works on the principle of change in resistance to magnetic field. There are two effects in this respect: Hall effect and anisotropic magnetoresistive (AMR) effects. Magnetoelastic material is the material whose shape changes with respect to magnetic field. Materials that retain residual magnetism without needing outside influence are called permanent magnetic materials. These materials retain their magnetism after being removed from a magnetic field. Some iron compounds or alnico when brought into the magnetic field, magnetize and retain their magnetism even after the removal of the field. The retentivity varies from material to material. Permanent magnets are very useful for other applications like actuation and data storing purposes.

Magnetic material based diodes and transistors are called magnetodiodes and magnetotransistors respectively, and can be used for magnetic field sensing applications. Magnetotransistors are more sensitive than magnetodiodes. Further integrated magnetotransistors can be designed to detect axial components of magnetic fields vector. Note that the magnetodiodes and magnetotransistors are microdevices rather than MEMS devices due to the reason that they do not have any movable parts. However, arguably they can be included under MEMS group of products. A true MEMS magnetic

sensor has a movable plate, which deflects in response to the magnetic field to be detected. There is an effect called MOKE (magneto-optic Kerr effect), which can be exploited to measure the pressure. When pressure is applied, the permeability of magnetic material changes. The permeability is measured by MOKE effect, which corresponds to a change in polarization state of light reflected from the magnetic material.

Magnetic actuators work based on the principle of magnetic force exerted on a actuating element (cantilever, paddle, probe, tip, etc) developed because of applied magnetic field generated from either from a permanent magnet or electromagnet. There are several applications of magnetic actuators including optical switching in communication field and microsurgery device actuation in health science. One of the major applications of magnetic MEMS is found in the computing sector in terms of micromachined storage devices. A thin magnetic storage medium (film) integrated with magnetic-probe and associated passive element is a promising technology, which can store more than 2 GB of data on 2 cm^2 of die area.

This chapter describes all the fundamentals, design, construction, fabrication, methodology advantages of magnetic sensors, actuators and storage devices in proper order.

Points to Remember

- Magnetic sensors have long been in use. Primitive applications were for direction finding used in navigation.
- MagMEMS devices are continuously being developed to form miniature structures using MagMEMS materials for a variety of purposes.
- The magnetic actuators are used in microsurgery applications, data storage application, and so on.
- MagMEMS materials are classified into three categories, Magnetoresistive materials, Magnetostrictive or Magnetoelastic materials and Permanent magnet materials.
- The resistivity of magnetoresistive material changes as the function of the external magnetic field.
- There exist some materials which change their shape when placed near a magnet.
- In a permanent magnet, however, the magnetic fields of the individual atom are aligned in one preferred direction, giving rise to a net magnetic field.
- Magnetic sensors perform extraordinarily in the magnetic field of below 1 gauss. The sensor can help in detecting ferrous objects, which disturb the magnetic field of the earth.
- There are primarily two physical effects, where the resistance is dependent in the presence of magnetic field. These are called Hall effect and Anisotropic Magnetoresistive (AMR) effect.
- Hall effect is pragmatic in all metals, but for the fabrication of the transducers, n-type semiconductor materials are preferred because it is easier to dope a controllable amount of charge carriers into the semiconductor material determining the application domain of the transducer.
- Magnetodiodes and magnetotransistors are two semiconductor-based devices used for the detection of magnetic field.
- The magnetodiode is made up of semiconductor p-i-n diode.
- The magnetotransistor possesses very large sensitivity compared to magnetodiode.
- While in operation the carriers are injected because of forward biasing of emitter-base junction. The operation of magnetotransistor is based on collector current detection.
- Magnetotransistor exhibits a linear magnetic response and can detect slowly or non-alternating signals.
- MEMS magnetic sensor detects the magnetic field vector by measuring the amplitude of a mechanical Lorentz force. The MEMS device consists of a current loop on an SiO_2 plate. Amplitude is detected with a polysilicon piezoresistor Wheatstone bridge.

- Sensitivity of a device plays a vital role in determining its capability and hence performance. This is a measure of ratio of change in transducer output to a change in the value of the measurand.
- Pressure can be measured using magnetic material. The principle of operation is that when pressure is applied the permeability of the magnetoelastic materials is changed. The permeability can be monitored utilizing the magneto-optic Kerr effect.
- MOKE corresponds to a change in the intensity or polarization state of light reflected from a magnetic material.
- Magnetic actuators work based on the principle of magnetic force exerted on a actuating element (cantilever, paddle, probe, tip, etc.) developed because of applied magnetic field generated from either a permanent magnet or electromagnet.
- A switch is a device consisting of a mechanical, electrical, electronic or optical element for making or breaking or changing the connections in a circuit.
- The magnetic actuation for switch application is called latching.
- The mechanical structure of the switch is that it has a silicon plate. Four 35 µm wide and 20 µm thick cantilever beams support the plate. Bonding pads are essentially required at the base to hold the beams with the substrate. The electromagnetic actuating component is a planner coil.
- Microactuators can be driven in both directions. Then these actuators are called bidirectional actuators. There has been a growing interest in the realization of bidirectionally driven microactuators in the areas of optical communication and computing devices such as scanners, etc.
- The remote magnetic actuator with feedback consists of two basic units: actuation unit and feedback unit.
- Magnetic-probe based storage device utilizing MEMS technology is a promising technology, which can store more than 2 GB of data on 2 cm^2 of die area.
- MEMS magnetic probe based storage device can access gigabytes of high-density, nonvolatile magnetic storage data and approximately more than 6 times faster than that of a modern disk.
- Data are written (stored) and retrieved by positioning and moving the media sled over the probe tip array. The media sled is a thin film magnetic media and the tips are permanent magnetic tips.
- Writing is carried out with the probe tip by simply modulating the spacing between the tip and the media. Reading of data is accomplished by knowing the magnetic state of the media, which is achieved by sensing the force that the media exerts on the permanent magnet head.
- A column in the rectangular region is called *tip track*. A tip track has M bits and is organized into several vertical *tip sectors*.
- A tip track is divided into multiple *logical sectors*.
- Each bit can be represented by something called a *triple* such as <*x*, *y*, tip>. Here <*x*, *y*> represent the coordinates of the bit of interest and the <tip> signifies the probe tip within that addressable region.

Exercises

1. Classify the MagMEMS materials according to their properties.
2. Explain the principle of magnetoresistive, magnetostrictive and hard magnetic materials.
3. Draw the hysteresis curve and explain its nature.
4. With a suitable diagram explain the principle of operation of a magnetoresistive sensor. In particular, discuss the operation of Hall and AMR effect magnetoresistive sensor.
5. Write notes on the following microdevices.
   (a) Magnetodiodes
   (b) Magnetotransistors

6. What do you mean by integrated magnetotransistor? What important design criteria are considered while fabricating the integrated magnetotransistor? List out the advantages of integrated design.
7. Justify whether or not magetodiodes and magnetotransistors are MEMS devices.
8. Discuss the principle of operation and construction of a typical microplate type MEMS magnetic sensor that can detect the magnetic field or current in a circuit. What do you mean by sensitivity of the magnetic sensor? Plot the sensitivity curve and explain.
9. What is the expansion of MOKE? Explain how a MOKE device works? Plot the typical pressure and permeability curve and explain how it relates to the act of sensing?
10. Define a MagMEMS actuator.
11. Discuss the constructional details and operational feature of a typical 2×2 optical switch that should be operated on magnetic based actuation.
12. Distinguish between unidirectional and bidirectional magnetic actuator. Explain the constructional details and principle of operation of a cantilever based bidirectional magnetic actuator. List out the important process sequence with regard to the fabrication of magnetic actuator. Consider a two-cantilever based magnetic actuator. Write down the expression for maximum angular displacement and maximum vertical deflection occurring at the free of the cantilever. Define the terms used in this expression.
13. Write notes on the following actuators:
    (a) Feedback circuit integrated magnetic actuator
    (b) Large force reluctance magnetic actuator
14. Comprehensively discuss the principle of operation of magnetic probe based storage device.

Chapter 9

Radio Frequency (RF) MEMS

Objectives

The objective of this chapter is to study the following.

- Introduction to radio frequency (RF) communication system
- The applications of MEMS in radio frequency (RF) communication
- RF inductors: Planner and solenoid inductors along with the fabrication method
- Varactors and fabrication methods
- RF tuners, Filters and Resonators
- Switches
- Phase shifters: Linear and distributed
- Tuning principle and phase array antenna

9.1 INTRODUCTION

Wireless communications have existed for a long time. New wireless communication systems are being developed more rapidly than ever. Wireless technology utilizes RF (Radio Frequency) signal, which is an electromagnetic (EM) signal. Radio frequency operates in the range 9 kHz to 300 GHz.

Radio frequency microelectromechanical systems (RF MEMS) is a field that is concerned with development of micromachined devices such as filters, oscillators, resonators, and switches, aimed at high frequency (~1 MHz to 60 GHz) communication applications. Fundamentally, the filters, oscillators and switches are composed of passive elements such as inductor and capacitor. Resonator, on the other hand, is a hollow chamber (cavity structure). By manipulating the value of inductance, capacitance and cavity dimension filtering, oscillation and resonance can be achieved, respectively. RF MEMS are the product of material science, circuit technology, mechanical engineering and communication methods. Innovative designs have made it possible to fabricate devices operating at wide frequency ranges. This enables RF MEMS to deliver integrated RF components on the same wafer.

RF MEMS can be used for achieving:

- Transmission and reception
- VCO tuning
- RF band select filters
- Intermediate Frequency (IF) filtering

- Time delay for phased-arrays
- Variable Delay Lines (VDL)
- Reconfigurable antennas design

9.2 REVIEW OF RF-BASED COMMUNICATION SYSTEMS

All communication systems consist of three main building blocks, namely transmitter, receiver and communication media. The transmitter transmits the baseband original signal by adopting appropriate modulation techniques[1]. The modulated signal is transmitted to the channel through an *impedance-matching* unit. The impedance is defined as the opposition that a circuit offers to the modulated signal or any other varying current at a particular frequency. The receiver receives the modulated signal from the channel and demodulates it to get back the original signal. There exist various types of channels, including optical fiber, conducting wire, cable and air. When air is used as the channel, antennas are required at both the ends; one at the transmitting end to transmit the modulated RF signal and another at the receiving end to receive the RF signal. In this case antennas are impedance matching units. Such a communication is defined as wireless communication system and the frequency of operation of the system falls under RF range (Table 9.1).

Table 9.1 *RF frequency of operation*

Frequency range	*Designation*	*Wavelengths*
9–30 kHz	Very Low Frequency (VLF)	33 – 10 km
30–300 kHz	Low Frequency (LF)	10 – 1 km
300 kHz–3 MHz	Medium Frequency (MF)	1 km – 100 m
3–30 MHz	High Frequency (HF)	100 – 10 m
30–300 MHz	Very High Frequency (VHF)	10 – 1 m
300 MHz–3 GHz	Ultra High Frequency (UHF)	1 m – 100 mm
3–30 GHz	Super High Frequency (SHF)	100 – 10 mm
30–300 GHz	Extremely High Frequency (EHF)	10 – 1 mm

The basis of communication is to deal with the transmission of spectral power of the desired frequency component or band (group of consecutive frequency components) from one point to another in an effective way. The transportation of power is accomplished by the use of a series of

[1] *Modulation* is a process of changing the parameters of a high frequency signal called *carrier signal*, with respect to the intensity of a given weak signal called *original baseband* signal or *modulating* signal. The high frequency signal is usually a sinusoidal signal. The parameters are simply the amplitude, frequency and phase. The modulation is essential in communication systems, where a weak signal is transmitted by the use of a carrier signal. There are many forms of modulation such as Amplitude Modulation (AM), Frequency Modulation (FM), Phase Modulation (PM). When a high frequency signal has amplitude varied in response to the amplitude of a low-frequency weak signal the modulation is said to be AM. When the frequency of the high frequency signal is varied in accordance with the amplitude of the low-frequency weak signal the modulation is said to be FM. Similar definition can also be given for PM. The process of recovery of original baseband signal from the modulated signal is called *demodulation*. There exist other modulation technique, called pulse modulation, which includes Pulse Amplitude Modulation (PAM) and Pulse Width Modulation (PWM) and Pulse Code Modulation (PCM). PCM is a digital modulation technique.

communication modules, whose roles are very explicit to achieve the final goal, i.e. communication. The block diagram of a typical communication system accommodating various modules is illustrated in Fig. 9.1.

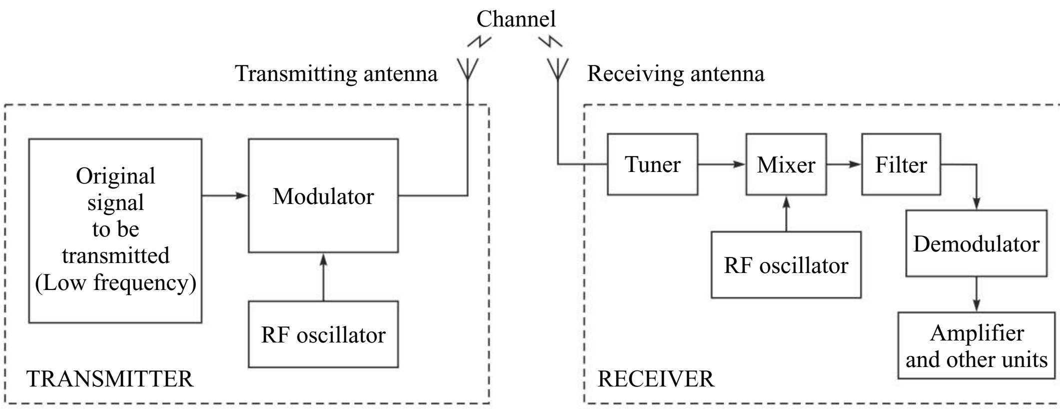


Fig. 9.1 *A typical communication system*

As the scope of this book does not permit detailed explanations of the principle of operation of each module and hence the communication system, students are advised to gather introductory knowledge on this topic from other fundamental books (Recommended book: *Principle of Communication Systems* by Taub and Schilling – Tata McGraw-Hill). Now, coming to our original discussion, it is important to note that the building blocks that are used in RF communication applications are called RF modules. Important RF modules are listed in Table 9.2.

Table 9.2 *List of RF modules*

List-1	*List-2*
Amplifier	Tuner
Modulator	Resonator (RF Oscillators)
Mixer	Coupler (also called switch)
Filter	Phase shifter
Isolator	
Phase Locked Loop (PLL)[2]	

Modules under List-1 are very common in other non-RF applications, but modules under List-2 draw attention as far as RF MEMS are concerned. Modules under List-1 primarily deal with electrical signals for which they are designed by the use of discrete electronic components or Integrated Circuits (IC). On the other hand, the modules enlisted under List-2 deal the RF signals. A brief description follows.

9.2.1 Tuners

Tuner is a frequency selective module that allows a band of frequencies to pass through. You have experience in tuning a broadcast channel in your pocket radio receiver. The signal available in the air is a composite signal that consists of many channels with different band of frequencies. You are tuning to select a particular band of frequencies, i.e. channel. When the tuner frequency band matches with the desired frequency band out of many frequency bands available in the air, then the desired frequency band is said to be selected and as a result of which you are able to listen particular channel. This is as

[2] PLL can be used for demodulation of FM signal

good as saying that the power within the signal of a desired frequency band available in the air could be transferred to the pocket receiver (in this example) if it is tuned properly. The transformation of maximum power will occur only when the input impedance of the tuner circuit equals the impedance of the air channel. Generalizing the above, it can be concluded that when the output impedance of the previous circuit matches the input impedance of the next circuit then maximum power will be transferred from the former to the latter one. However, if the two impedances do not match and there is a requirement to transfer maximum power from one stage to the other then we need to design the impedance matching circuit or system in between them in order to achieve the objective. Note that the transmitting antennas are the impedance matching systems, since they transfer maximum power from the transmitter (previous stage) to the air (next stage). Similar concept can be applied to receiving antennas. The applications of impedance matching circuits are found everywhere in a communication system. Tuners are also impedance matching circuits since they are used to transfer power.

9.2.2 Resonator

Resonator, being a hollow chamber, is a special kind of module. The dimension of the hollow chamber plays important role in making resonance. In fact, by varying the dimension of the chamber, resonance can be achieved. This is again called tuning. When tuned properly the resonator can produce RF or MF (microwave frequency) signal at the output. A very precise stable oscillatory signal can be generated through resonance. This is why resonators are also called oscillators. It can be summarized that the chambers whose dimensions allow the resonant oscillation of electromagnetic or acoustic waves are referred to as resonator or cavity oscillators. The resonators are two port devices.

9.2.3 Switch

Switches were introduced in Chapter 4. Sometimes, switches are called couplers. This chapter deals with MEMS switches. MEMS switches have demonstrated outstanding performance from 0–100 GHz compared with their solid-state counterparts. Two types of switches are available; contact and non-contact types.

9.2.4 Phase Shifter

A phase shifter is also a two-port device that alters the phase of its output signal in response to an external control signal. The phase shifting device is used to virtually flip an antenna or arrays of antennas in relation to producing propagation angles of a desired HF (high frequency) path. More descriptions will be provided later.

9.3 RF MEMS

Clearly, the RF designs consist of two types of modules (modules under List-1 and List-2) due to the reason that it was difficult to fabricate inductors and capacitors in a single platform. However, because of the advent of MEMS technology, the RF design can be integrated in one cabinet in order to achieve greater integrability and modularity. Dedicated MEMS devices, used for RF applications, are called RF MEMS. RF MEMS technology is a high-quality, three-dimensional, microscale structure that is manufactured using micromachining batch fabrication techniques.

9.3.1 Application Areas

The application areas are aerospace, defense, instrumentation, commercial, and so on. The RF MEMS can be used where there is a requirement of:

- RF communication (Frequency of operation is tens to hundreds of KHz)
- Microwave communication (Frequency of operation is hundreds of MHz)
- Global positioning systems (GPS)
- Broadband wireless access and wireless data link

In essence, RF based communication systems are used in applications, such as mobile communication, mobile robots, navigation and broadcasting.

9.3.2 Advantages

Traditionally, the RF modules are realized using PIN (p-intrinsic-n) diode, IMPATT diode (***IMP***act ionization ***A***valanche ***T***ransit-***T***ime), waveguide and coaxial switches. Although, these devices show low insertion loss and good isolation, they are power-hungry, slow, and unreliable for long-life applications. The relative advantages of RF MEMS technology over the traditional non-MEMS based designs mentioned above are listed below.

- Highly-integrated and portable
- Very high quality factor (Q-factor)
- Linearity
- Low power consumption
- High switching speed
- Low signal level operation
- Low allowable losses (insertion loss)
- Availability, low cost, and ruggedness.

Table 9.3 compares the important performance parameters of MEMS- and non-MEMS-based RF devices.

Table 9.3 *A comparison of performance parameters of MEMS-based and non-MEMS-based RF devices*

PARAMETER	*RF MEMS*	*PIN-DIODE*	*FET*
Voltage	20–80	± 3–5	3–5
Current (mA)	0	0–20	0
Power Consumption (mW)	0.5–1	5–100	–.5–0.1
Switching	1–300 ms	1–100 ns	1–100 ns
C_{up} (series) (fF)	1–6	40–80	70–140
R_s (series) (W)	0.5–2	2–4	4–6
Capacitance Ratio	40–500	10	n/a
Cutoff Freq. (THz)	20–80	1–4	0.5–2
Isolation (1 – 10 GHz)	Very high	High	Medium
Isolation (10 – 40 GHz)	Very high	Medium	Low
Isolation (60 – 10 GHz)	High	Medium	None

Loss (1 – 100 GHz) (dB)	0.05 – 0.2	0.3 – 1.2	0.4 – 2.5
Power Handling (W)	<1	<10	<10
3rd order Int. (dBm)	+66 – 80	+27 – 45	+27 - 45

(*Source:* NASA's EEE Parts & Packaging Program (NEPP))

9.3.3 Design Scenarios

MEMS conformant RF module design schematics are based on four fundamental architectural principles such as,

- Inductor-based design
- Varactor-based design
- Resonator-based design
- Switch-based design

A MEMS tuner can be designed using MEMS inductor or MEMS capacitor or combination of these two fundamental elements. Similarly an MEMS oscillator for both modulation and demodulation purpose can be designed utilizing resonator-based design approach. MEMS switches are very important RF circuit components and used for transmission of RF power from input port to the output port. All these modules are very fundamental, as far as MEMS based design of RF communication systems are concerned.

9.4 MEMS INDUCTORS

The purpose-made conductor coils are called inductors. The behavior of inductor in response to an alternating (ac) signal is that it opposes the input current change. Inductors are the basic building blocks of all types of oscillator, delay and actuating circuits. RF inductors are those which are used at RF frequency range. RF inductors have already been designed and in use based on MEMS technology.

An inductor is a passive element that stores energy in a magnetic field, typically by combining the effects of many loops of current. When electromagnetic RF signal flows through RF inductor an electromagnetic field is established and it changes in accordance with the input signal. The changing electromagnetic field causes an induced voltage in a direction in opposite to the flow of input signal. This property is referred to as *inductance*. The magnetic flux thus established is proportional to the strength of the input signal. There are actually four types of inductors, namely straight, spiral, solenoid and toroidal. All these types are presented in Fig. 9.2(a). Out of these spiral and solenoid types are popular because of high Q (described later). With regard to spiral and solenoid type inductor, the strength of the induced voltage will increase if the conductor coil contains more number of turns.

9.4.1 Planer Inductor

RF microinductors are of two types, namely solenoid type inductor and planer coil inductor. The solenoids are conductor coils. The application areas of planer coil inductors are seen to be more as compared to the latter type. The planer inductors are either rectangular or circular. Figure 9.2 shows SEM picture of a rectangular planer RF MEMS inductor. The equivalent circuit of the planer RF inductors is given in Fig. 9.2. Here,

L_s = Low frequency inductance,
R_s = Series resistance,
C_s = Capacitance between the different wings of the inductor,
C_1 = Capacitance in the oxide layer between the coil and the silicon substrate,

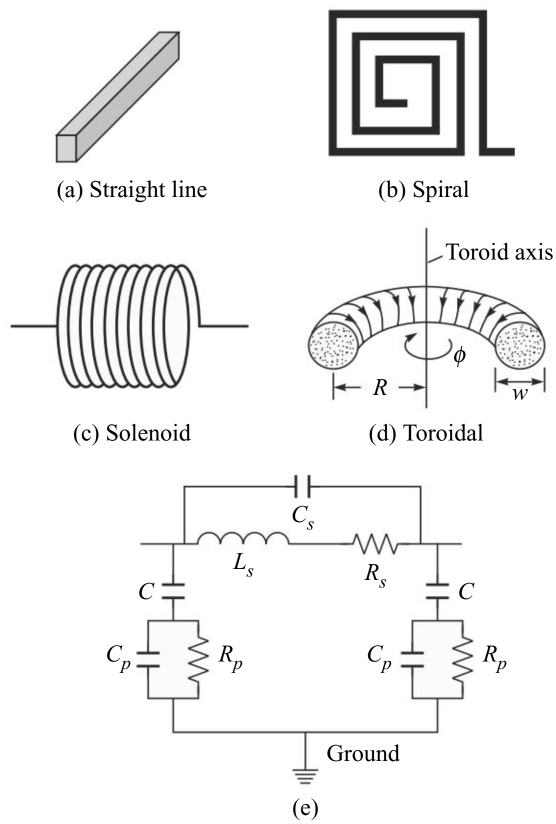


Fig. 9.2 *(a–d) Schematic diagram of various types of inductors* (*Courtesy:* Huang and Cao, Dept. EECS, University of California, Berkeley, USA), *(e) RLC model of planer RF inductor* (Rebeiz, Wiley-Interscience, 2003).

C_p = Capacitance between the coil and the ground through the silicon substrate,
R_p = Eddy current losses in the substrate.

The quality Q of an inductor is the measure of inductiveness of the coil. In other words, Q expresses the merit of the inductor. A higher value of Q is desirable. Although it is possible to achieve Q in the order of 900 or so, in practice, an inductor has Q values less than 100 when used within its operational RF frequency ranges. Mathematically, the Q of an inductor is given by,

$$Q = \frac{wL_s}{R_s} \tag{9.1}$$

where Q is the quality of the inductor, L_s and R_s are the inductance and resistance of the inductor respectively, with the assumption that the inductor is operating at low frequency. At low frequency the effect of capacitance can be neglected. From Eq. 9.1 it is observed that as $R_s \rightarrow 0$, $Q \rightarrow \infty$. So the resistance of the coil has to be extremely low in order to obtain high quality factor. The Q values with respect to the frequency of operation of typical RF inductors are shown in Fig. 9.3 and Table 9.4.

Table 9.4 *Typical Q-factor of RF inductor*

Inductor design values (mH)	*Q at 2 GHz*	*Q Maximum*	*Frequency at which Q is Maximum*	*L_s (mH)*	*Resistance at 2 GHz (Ohm)*
1.5	41.0	45.0	4.85	1.55	0.515
2.5	75.0	78.5	1.95	2.99	0.530
4.5	60.0	65.0	1.85	4.50	1.260
12.5	19.0	25.5	1.40	13.15	6.655

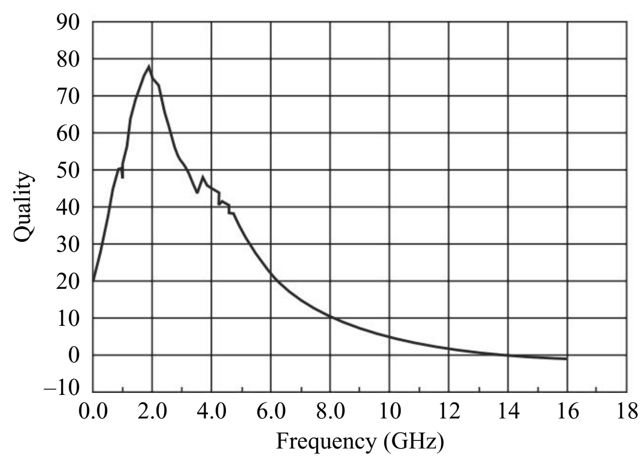


Fig. 9.3 *Q-values of typical RF inductor with respect to the frequency of operation*

9.4.2 Fabrication Process

Various micromachined processes are involved in manufacturing the planer inductors. Accordingly we have primarily three phases of developing the planer inductors; (a) micromachining using thick metal layers; (b) substrate etching; and (c) assembly.

Detail fabrication process is illustrated in Fig. 9.4. The substrate is a silicon wafer with approximately 0.6 μm thick nitride layer on top. The fabrication follows the following steps.

- Half a μm-thick silicon oxide layer is deposited and patterned to serve as the sacrificial layer (Fig. 9.4(a)).
- Another half of a μm-thick gold layer is deposited onto the substrate (i.e. on the top of the sacrificial layer) (Fig. 9.4(a)).
- The gold layer is patterned to make the bottom conductor (coil) of the inductor. Gold is a ductile material with high conductivity. It is an ideal plastic deformation material. The plastic deformation property is essential in making planer type RF inductors. Refer Fig. 9.4(b).
- Approximately a quarter μm-thick of CYTOP film is spiraled onto the gold layer and patterned. This is an amorphous fluorocarbon polymer film that essentially makes the dielectric spacer. The electrical properties of the CYTOP are similar to the Polyimide and Telfon. However, it has proved to be better adhesive characteristics on metal surfaces and also improved chemical stability and temperature resistance (Fig. 9.4(c)).

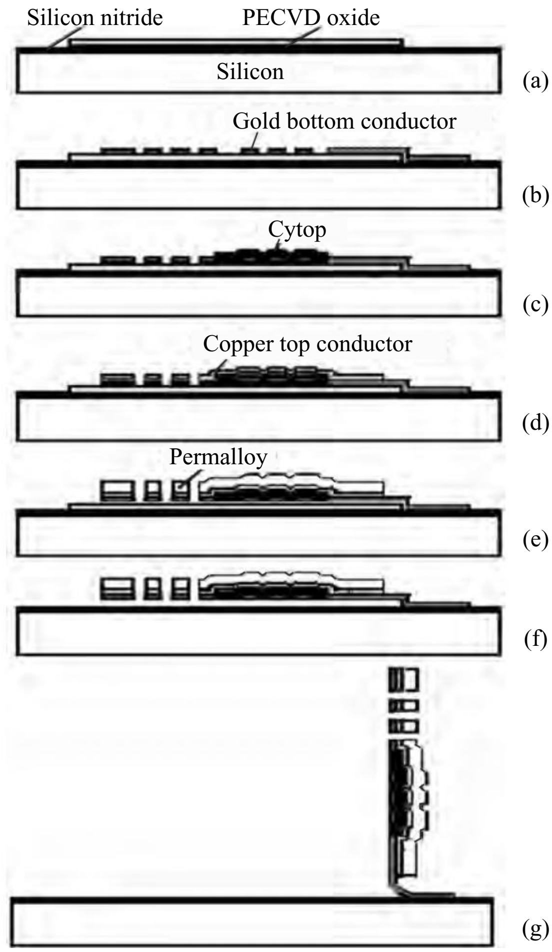


Fig. 9.4 *Fabrication process*

- A 1.5 µm thick copper layer is deposited and patterned. This helps in preparing the upper conductor of the inductor. Copper is usually selected mainly due to the fact that it has high conductivity and the processing is compatible with the gold bottom conductor during the fabrication (Fig. 9.4(d)).
- Then a 5 µm thick Permalloy layer has to be electroplated onto the copper and gold surface (Fig. 9.4(e)).
- The oxide sacrificial layer is finally etched (Fig. 9.4(f)).
- The structure is released from the substrate as shown in Fig. 9.4(g). Figure 9.5 shows an SEM photograph of planer MEMS inductor before and after the release.

9.4.3 Solenoid Inductor

Some RF applications require current to be confined to the surface of the conductor. One such building block is the antenna. RF antennas are of various types. One type is a solenoid structure. Solenoid

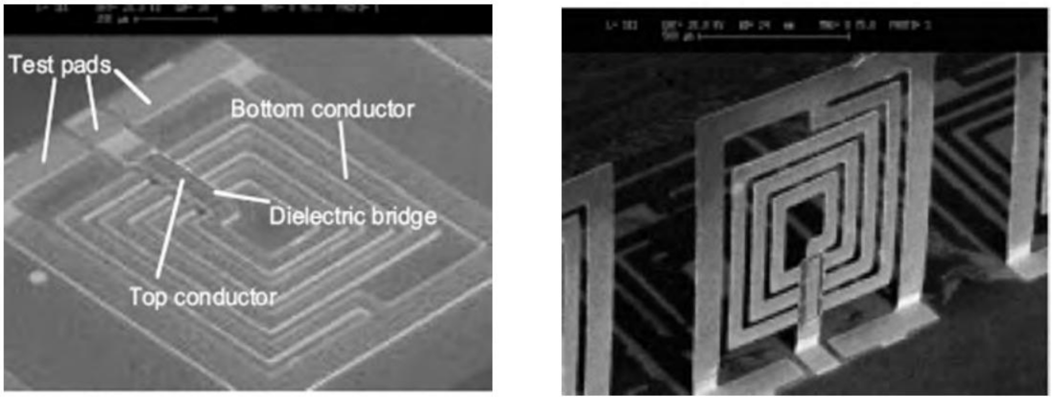
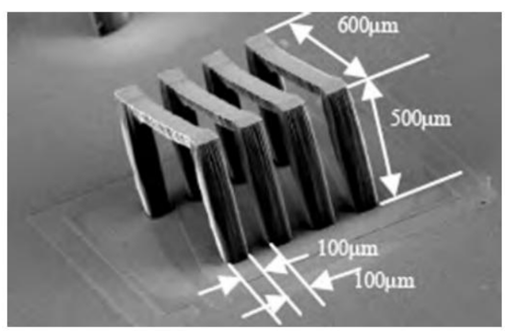
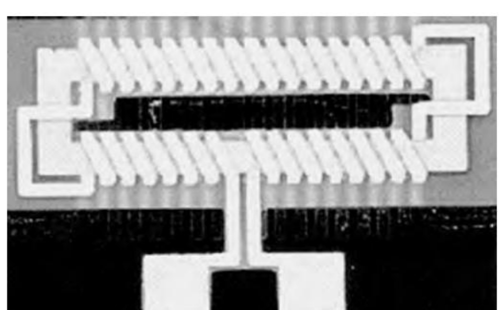


Fig. 9.5 *SEM photographs of planer MEMS inductor* (*Source:* Chen et al, *Agere Systems*, U of Illinois at Urbana, U of California)

inductor can thus be used for tuning applications. In a receiving antenna, the time variant radiated electromagnetic current appears at the surface of the solenoid structure. Figure 9.6 shows typical solenoid type inductors. The solenoid is a longitudinal device composed of conductor coils usually from a single conductor and a spiral in structure. However, the physical structure as far as MEMS solenoids are considered is slightly different. For example, the coils could be circular, rectangular, hybrid or any



(a)



(b)

Fig. 9.6 *(a) A rectangular microsolenoid, (b) A spiral solenoid (Dimension—4 mm in length, 1 mm in width, and 0.13 mm in height—Source:* Jae Park and Mark G. Allen, Georgia Inst. of Technology*)*

other structure meeting the RF characteristics and requirements. The internal space/block of the coil is called core. The core is usually made from another material. Core-less solenoids are called air-core solenoids. Typical single turn inductor (solenoid) is approximately 500 μm in height and 600 μm wide. With air-core the inductor can show a Q-factor of 60–85 and an inductance of 1 to 2 nH at 2.5 GHz. In solenoids current is confined to the outermost portions of conductors due to the skin effect.

Attempts have been made to design semiconductor-based inductors. But the known limitation of semiconductor based on-chip spiral inductors is their low Q because of lossy substrate materials. A MEMS inductor on the other hand overcomes the drawback because of increased spiral height above the substrate. The increase in height with more air surrounding reduces the stray capacitance and dielectric losses. MEMS inductors have both increased inductance values and greatly improved Q over traditional inductors.

9.5 VARACTORS

The short form of variable capacitor is called varactor. Varactor is an active[3] device whose capacitance value is varied by some means. A varactor is also called *varicap*. The varactors are classified into two groups based upon the technology using which they are manufactured: accordingly they are semiconductor varactors (also called solid-state varactors) and MEMS varactors. MEMS varactor is a recent development.

Semiconductor varactors have long been in use not only in the RF communication systems but also in many other applications, such as instrumentation and control. Traditionally, varactor has been considered as a semiconductor diode with the properties of a voltage-dependent capacitor. Semiconductor varactor is a junction diode that exploits the depletion layer in terms of a parallel plate capacitor, whose capacitance is controlled by applying reverse voltage across the junction of the diode. The reverse voltage controls the width of the depletion layer, which in turn controls the value of capacitance because the capacitance across the junction is inversely proportional to the width of the layer.

Advantage: MEMS varactor overcomes some of the drawbacks encountered in semiconductor-based device. The relative merits of MEMS varactors are,

- Low power consumption
- High quality factor
- Tolerance for high voltage swing
- Low harmonic distortion
- Large tuning range

Low power consumption and high quality factor is always desirable. Changing the biasing voltage varies the capacitance of the varactor. The varactors can tolerate large voltage range. Hence it can facilitate large tuning range. The varactors are useful for designing VCO (Voltage controlled Oscillator), FM (Frequency Modulator) modulators and demodulators and are used to design tuning circuits. The MEMS varactor, when coupled with an inductor, forms an LC tank circuit (oscillator) offering high Q-factor and better performance. A typical VCO implementation would include the MEMS varactor coupled with an inductor in a feedback loop sustained by a negative resistance amplifier.

[3] In electrical engineering and electronics usually resistors, capacitors and inductors are called passive components. Active devices are based on semiconductor technology. These devices are based on the p-n junction. The diode (can rectify the input signal) and transistor (can amplify the input signal) are based on single and double junctions, respectively. A diode, although it does not amplify, still is considered as an active device. However, the categorization of passive and active elements in MEMS domain is somewhat different. Refer to the descriptions presented in Chapter 4.

Theory: The varactor is constructed by mechanically suspending a metal plate in the air above a fixed metal plate. In this way a parallel plate capacitor (PPC) is formed. A parallel plate MEMS varactor is illustrated in Fig. 9.7. The suspended plate is called top plate and the fixed plate is called bottom plate of

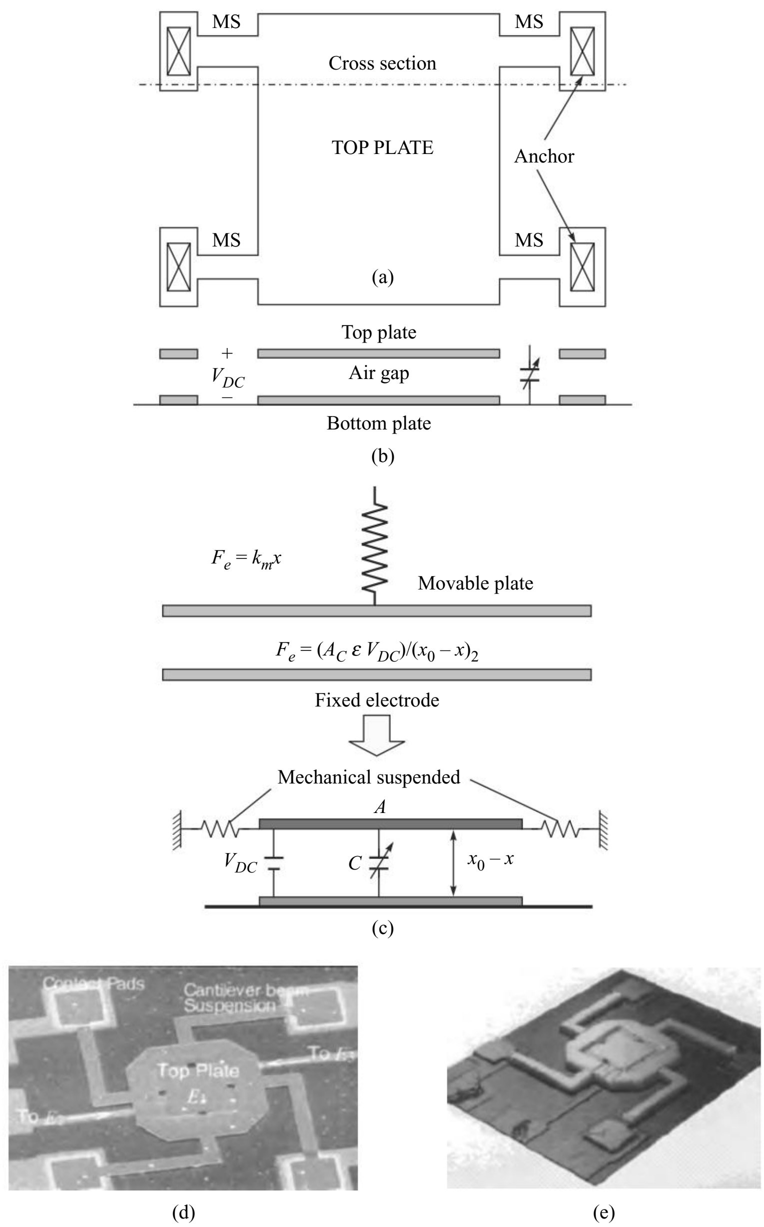


Fig. 9.7 *A schematic parallel plate RF MEMS varactor (a) Top view (MS: Mechanical Suspension), (b) Cross-sectional view; (c) Showing electrostatic actuation of the plates; (d) The photograph of a real MEMS varactor; (e) Another MEMS varactor (Source:* McCorquodale et. al. *IEEE MTT-S Digest*, IFTU-54, 9, 2003*)*

the PPC. The top plate is movable. Four anchors are necessary to achieve suspension (Fig. 9.7(a)). The anchor in conjunction with associated elements is called mechanical suspension network (MSN). The MSN eventually provides support for the top plate.

The nominal capacitance value is determined by the device geometry and the nominal gap between the plates. Capacitance in a parallel plate capacitor is given by,

$$C = \frac{\varepsilon\varepsilon_0 A}{x_0} \tag{9.2}$$

where, C is called the un-modulated or resting-state capacitance, ε is the dielectric constant of the material present in between the two plates, ε_o is the dielectric constant of the free space (air), A is the overlapping effective area of the plates and x_o is the distance of separation. By applying a positive DC voltage (electrostatic force) across the plates, the movable top plate exerts a force and gets deflected by some distance. The applied DC voltage is called tuning voltage because the varactor is primarily used for tuning purpose. The deflection causes capacitance to change. Therefore, the above equation can be reduced to,

$$C = \frac{\varepsilon A}{x_0 - x} \tag{9.3}$$

where, x_o is the nominal distance between the plates, and x is the displacement due to force cased by the applied DC voltage. Thus, by changing the distance between the plates the capacitance can be changed. If this varactor is a circuit element of the tuning module the change in capacitance cause the frequency to change since, $\Delta C \; \alpha \; \Delta f$. MEMS varactor closely resembles a traditional variable capacitor. Besides air, inert gas can also be used as the dielectric in order to obtain very high Q factor. A desirable capacitance value can be obtained by using arrays of identical smaller capacitors. They are then called *gang varactors*.

The equation governing the electrostatic force generated between the plates by the applied tuning voltage is expressed by considering the energy stored in the plates.

$$F_e = \frac{\partial \xi}{\partial x} = \frac{1}{2}\frac{\partial C}{\partial x}V_{CD}^2 \tag{9.4}$$

$$= \frac{CV_{CD}^2}{2(x_0 - x)}$$

where, ξ is the energy stored in the plates. Since, F_e equals to $k_m x$, where k_m is the mechanical spring constant associated with the plate-anchor-suspension arrangement, the DC tuning voltage can be derived from the above equations as,

$$V_{CD} = \sqrt{\frac{2k_m x(x_0 - x)^2}{\varepsilon A}} \tag{9.5}$$

Cantilever beams or fixed-fixed beams are sometimes fabricated in place of plates as shown in Fig. 9.8. Here, however, the mechanical spring constants are,

$$k_{\text{cont}} = \frac{2}{3}EW\left(\frac{t}{L}\right)^3 \tag{9.6}$$

$$k_{\text{fixed}} = 32EW\left(\frac{t}{L}\right)^3 \tag{9.7}$$

where, k_{cant} is the spring constant of the cantilever, k_{fixed} is the spring constant of the beam, E is the Young's modulus of the material, and W, t, L are the width, thickness, and length of the suspended plate, respectively.

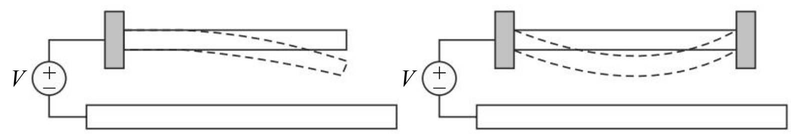


Fig. 9.8 *Cantilever beam and fixed-fixed beam capacitor*

Typical range of capacitance of MEMS varactors is 1–15 pF with actuation voltage up to 8V, which can provide Q approximately 100 in the range of frequency operation of 200 MHz to 500 MHz.

Fabrication: For varactor, the parallel-plate configuration using electrostatic actuation is most commonly used. Surface micromachining process is used to fabricate the varactor. Pyrex glass wafer of 1 mm thickness is taken as the substrate. The sacrificial layer is then developed. Thermally evaporated gold thin film is used as the material for the bottom plate. There may be more than one bottom plates. The top plate is made up of electroplated Permalloy. Permalloys are nickel-iron alloy. The thickness of the Permalloy deposition can be up to 200 µm. The Permalloy has the desirable characteristics of good smoothness and low stress. A brief description of the process is illustrated below.

- Initially a 0.5 µm-thick gold film is thermally evaporated and patterned to form the two fixed bottom plates E_2 and E_3 (Fig. 9.9(a)). In this phase the contact pads are developed. Contact pads are used for the purpose of anchoring the suspended top plate E_1.

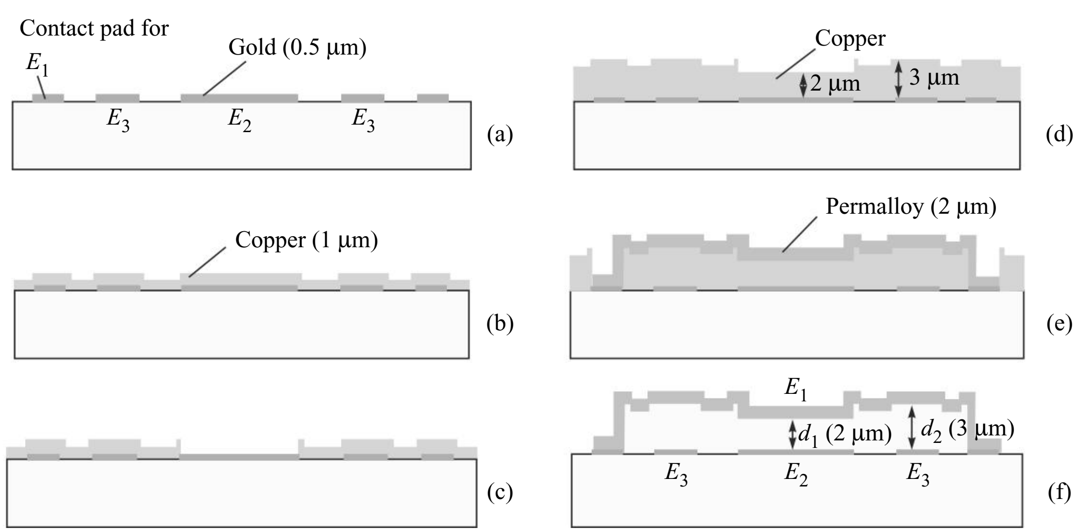


Fig. 9.9 *A schematic illustration of the fabrication process of MEMS varactor (Courtesy:* Interscience; *Source:* Zou et al., *Int. J RF and Microwave CAE 11,* John Wiley & Sons, Inc. 2001*)*

- Then a thickness of approximately 1 μm copper film is thermally evaporated and patterned.
- It is followed by another thermal evaporation to form another 2 μm thick copper film to make the variable-height sacrificial layer (Fig. 9.9(b-d)).
- In the subsequent stage 2 μm thick Permalloy is deposited by electroplating (Fig. 9.9(e)).
- Then the sacrificial copper layer is etched.
- Finally the fabricated device is placed in a dryer.

9.6 TUNER/FILTER

A simple capacitor-inductor circuit can act as the tuner for the RF receiver. It is usually connected to the antenna as shown in Fig. 9.10. The capacitor and inductor values are chosen to oscillate at one particular frequency. When the induced RF signal in the antenna matches with this frequency resonance occurs. The method of producing resoace is known as tuning and is achieved by the tuner circuit. The frequency at which tuning occurs is called resonant frequency. Usually, the capacitance value is varied by the use of a varactor.

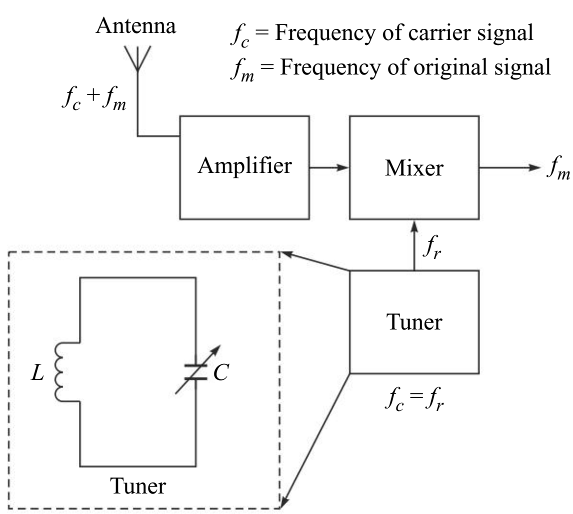


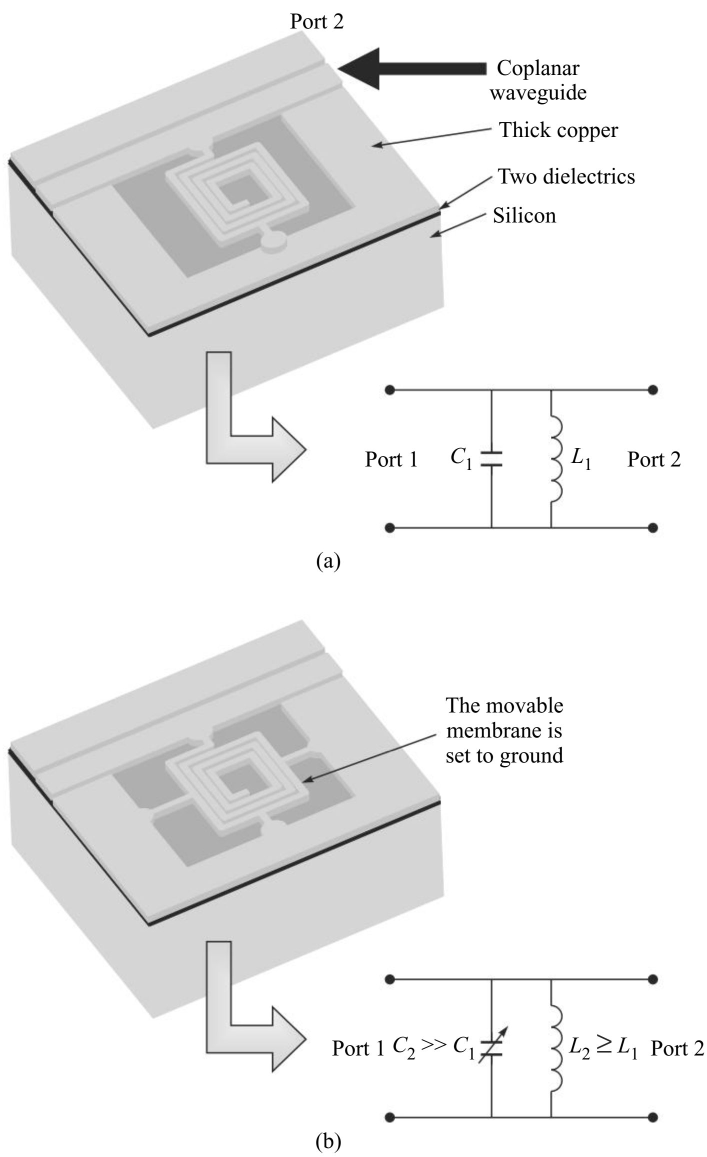
Fig. 9.10 *A schematic LC tuner*

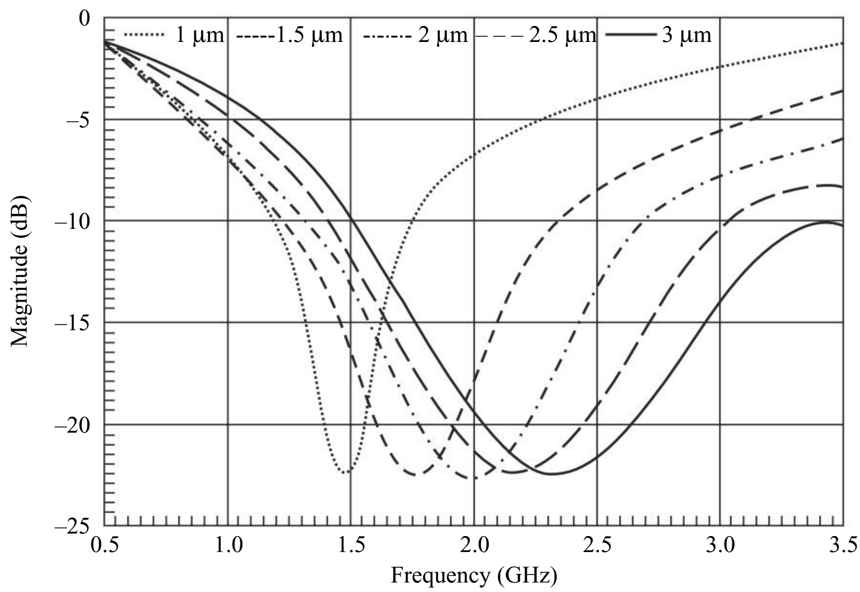
Figure 9.11(a) shows that the inductor is fabricated in thick copper and isolated from the underlying silicon substrate by a twin dielectric layer. The structure is equivalent to an inductor connected to a capacitor in parallel, constituting a filter or oscillator. On the top of the inductor a second conductive layer can be fabricated to act as a suspended membrane. Such a configuration can provide a large parallel capacitance, which can be varied by changing the gap between the spiral and the membrane (Fig. 9.11(b)). Standard MEMS actuation principles such as electrostatic or thermal actuation can be employed to vary the capacitance and hence the block can behave as a tuner. The frequency of oscillation is given by,

$$f_r = \frac{1}{2\pi\sqrt{L_1 C_1}} \tag{9.8}$$

where, f_r is the resonant frequency, L_1 and C_1 are the required value of inductance and capacitance to achieve resonance. Note that at resonance, the impedances of the capacitance and inductance cancels.

At resonance, ideally the maximum power is passed through, without any reflection. Reflection occurs if the value of inductance and capacitance are not chosen properly. The reflection is considered as the loss. The relationship between the frequencies versus loss for different capacitor gaps is illustrated in the Fig. 9.11(c). For instance, with capacitor gap 2 micrometer, the resonance frequency is 2 GHz and the loss is approximately –22.5 dB. The loss (reflection) increases with increase and decrease in frequency for a given capacitor gap.





(c)

Fig. 9.11 *Tunable MEMS technology-based LC resonator; (a) A single inductor layer can behave as a filter; (b) An inductor with variable capacitor behaves as Resonator; (c) Resonator performance vs. frequency for different capacitor gaps (Source:* Campbell (Ansoft Corporation Europe); *Tech On Line)*

The quality factor of the tuner can be defined in terms of frequency of operation. The Q is defined as the ratio of resonant frequency with the difference between the lower and upper cut off frequencies. Mathematically, it can be written as,

$$Q = \frac{f_r}{f_{+1/2} - f_{-1/2}} \tag{9.9}$$

where, $f_{+1/2}$ and $f_{-1/2}$ are the two half power frequencies. Q is dependent upon the design parameters including anchoring, material, pressure, temperature and surface quality. However, Qs tends to increase with volume. Therefore, determination of volume is important.

9.7 RESONATOR

In communication filtering of signal is common. Filtering implies that selection and rejection of desired and unwanted frequency components, respectively. For instance, if we require a signal, which should have a single frequency component or a group of frequency components then perhaps we need to feed the original signal, containing many frequency components to a filter, in order to extract the frequency component(s) of interest. A filter, which generates a single frequency signal, is sometimes referred to as oscillator. High precision single frequency signal is used for modulation and demodulation applications. LC tank circuit can generate a single frequency signal. However, resonators are popular where the need of high stability and selectivity arises. Further, resonators are useful for high frequency applications at which the LC oscillator does not have a reach. Resonator therefore, is a simply oscillator. Geometrically it is a hollow chamber. Figure 9.12 shows the schematic diagram of a cavity resonator. The dimension of the chamber plays an important role. The hollow space normally bounded by an

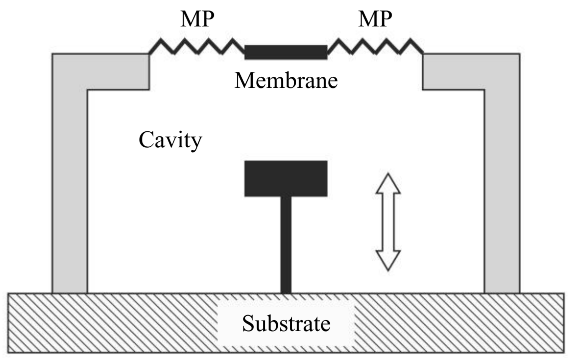


Fig. 9.12 *Schematic diagram of a cavity resonator (MP: Metal Plate)*

electrically conducting surface in which oscillating electromagnetic energy is stored. When the resonator is tuned it produces RF or MF (microwave frequency) signal. This is called resonance. The resonant frequency is determined by the geometry of the enclosure. Resonance is achieved by two methods,

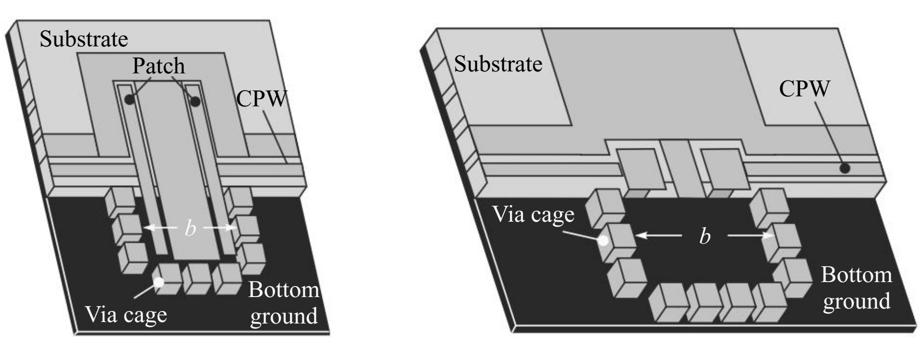
- Varying the input frequency
- Varying the dimension of the cavity

At resonance the input RF signal frequency and the resonance frequency match. Once the frequency matches the energy is said to be stored.

RF MEMS cavity resonator design parameters are not yet fully explored. There are various types of resonators as far as physical dimensions and shapes are concerned. Mostly the following types of resonators are encountered.

- Patch resonator and via resonator
- Microdisk resonator
- Bulkmode resonators

Patch resonator: The patch resonator (Fig. 9.13(a)) and via resonators (Fig. 9.13(b)) are three dimensional high resistivity silicon substrate filled cavity resonators. The backside of the substrate is metallized to form the bottom (ground) of the resonator. Via-hole fences realize its sidewalls. The top cover is formed by the metallized and patterned front side of the substrate. These two structures differ



(a) (b)

Fig. 9.13 *(a) The patch resonator, (b) The short-via resonator (Source:* Strohm et al., *IEEE MTT-S CDROM, 2002)*

with regard to excitation. The former one is called open-end patch resonator, the latter one is known as short-circuited via resonator. In the patch resonator, the electric field exists from the patch to the ground planes and the vias couple into the resonator. On the other hand, in the short-circuited via resonator, the magnetic field through the via excites the field in the resonator. In both the cases the minimum resonance frequency can be determined from the following formula.

$$f_r = C\sqrt{\frac{1}{L^2} + \frac{1}{b^2}} \tag{9.10}$$

where C is a constant depend on the properties of material used, L and b are the length and width of the cavity.

These resonators are built using co-planner waveguide (CPW) components. The resonators are fabricated directly in the substrate. The fabrication is performed on high-resistivity silicon wafers. The wafer has to be polished before the process. The wafer has to be cleaned in first place. After the cleaning process, the wafers are thermally oxidized to an oxide thickness of 50 nm. Then the oxide is removed at the portions where the via holes have to contact the resonator metallization top layer. The metallization layers of 20 nm Ti and 2200 nm Au are deposited and patterned. This process yields very smooth surfaces of the gold and very steep edges of the metallization layer. The via-holes are etched with an anisotropic deep silicon etch process. Vertical sidewalls are obtained with the dry etch process. Lastly the photoresist is removed in oxygen plasma. The via-holes and the backside of the wafer is metallized by sputtering.

Microdisk resonator: For high Q operations, microdisk resonators are used (Fig. 9.14(a)). E_i and E_o are the RF energy available at input and output port, respectively. E_d is the deflected energy. By

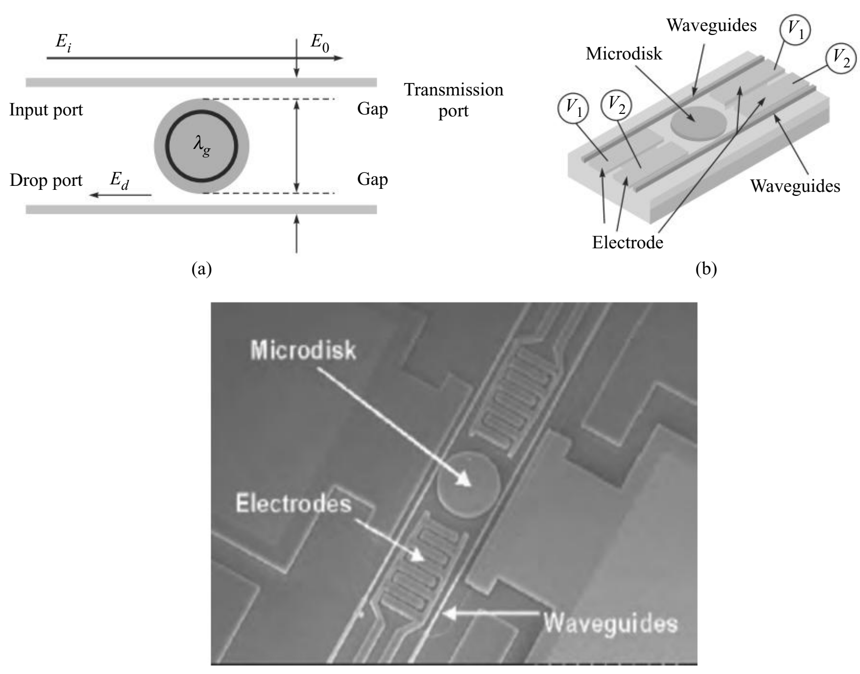


Fig. 9.14 *(a) A channel drop filter with microdisk resonator; (b) Schematic diagram of the variable-gap microdisk resonator. The waveguides near the disk are suspended; (c) The SEM picture of a fabricated device*

varying the gap spacing between the waveguide and the microdisks by a fraction of a micron more than ten orders of magnitudes change in the output transmission is effected. The device can also act as an wavelength-selective switch. The wavelength-selective switch is simply a tunable WDM (Wavelength Division, Multiplexing) component, commonly known as a dynamically add-drop filter.

The microdisk resonators are fabricated on a silicon-on-insulator (SOI) wafer. Two suspended waveguides are placed at proximity to the microdisk (Fig. 9.14(b)). The initial gap is about 0.9 μm wide, large enough to have no coupling between the waveguide and the microdisk. The suspended waveguides are pulled towards the microdisk by four electrostatic actuators. The radius of microdisk is 10 μm and the width of waveguides is about 0.7 μm.

***Bulkmode resonator*:** Bulkmode resonators are of two types, bulk longitudinal resonator (BLR) and bulk annular resonator (BAR). Both the types are useful for high frequency applications. They have large spring constant and small effective mass. Mostly BLR type devices (Fig. 9.15(a)) are popular because of its simpler design.

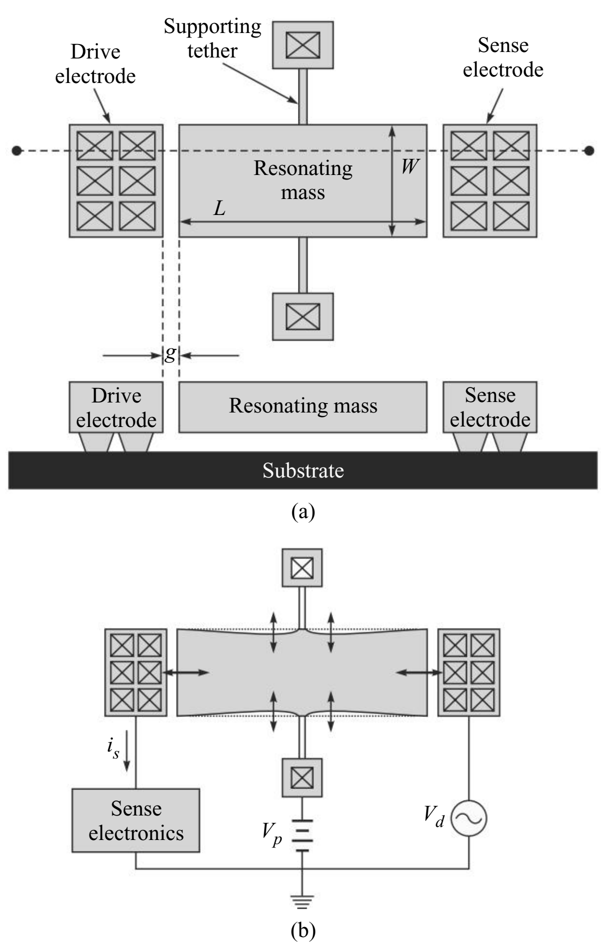


Fig. 9.15 *(a) Layout and cross-sectional views of the BLR; (b) Sample one-port biasing and operating scheme for the BLR. The dashed line represents the un-deformed BLR*

The BLR utilizes longitudinal and lateral vibrations, which apparently constitute as bulk vibrations. The applied RF signal should be near the resonant frequency. The drive electrode causes the resonator to extend and contract longitudinally as shown in the Fig. 9.15(b). The resonant frequency in this case is given by,

$$f_r = \frac{1}{2L}\sqrt{\frac{E}{\rho}} \tag{9.11}$$

where, E and ρ are the Young's modulus and mass density of the structural material, respectively. L is the length as shown. Note that the resonant frequency is independent of the width of the BLR. At resonance, the impedances of the capacitance and inductance cancel, leaving only the series resistance, R_{eq}. The resistance determines the performance of filters. A decrease in quality factor (Q) may occur due to the damping effect and the surface roughness of the resonator material.

9.8 CLARIFICATION OF TUNER, FILTER, RESONATOR

The reader should not be confused in understanding the term tuner, filter, oscillator and resonator. These modules essentially separate, generate and produce frequencies of interests. Although their internal operations and fundamental building blocks are observed to be similar, but their purposes are quite different. Since the common goal of these modules is to select some frequencies of interests, in the literature, these terms are used interchangeably. However, their objectives differ.

9.9 MEMS SWITCHES

Based upon the adopted technology, we have three varieties of RF switches namely, solid-state semiconductor switches such as FET and PIN diode; mechanical switches such as waveguide and coaxial; and the MEMS based switches. There are relative merits and demerits of the above three types of switches. Semiconductor switches provide faster switching speeds and are smaller in size and lower in weight. They have high insertion loss, power consumption and isolation characteristics. Waveguide and coaxial switches offer the benefits of low insertion loss, good isolation characteristics and high power handling capabilities. However, they are bulky, heavy and slow. The RF MEMS switches, on the other hand, are low in weight and small in size. Further, they consume less power. MEMS switches promise to combine the good properties of both mechanical and semiconductor switches. RF MEMS switches have been attracting users since they can replace many types of traditional semiconductor and mechanical RF components. Special attention is placed on switch performance at frequencies between 10–30 GHz. Typical switch parameters that the MEMS switch can provide for a cell phone applications can be seen from Table 9.5.

9.9.1 Types of MEMS Switches

Figure 9.16 shows the schematic diagram of a typical MEMS switch. MEMS switch is based on either capacitive coupling or physical contacts. The capacitive coupling is achieved by designing coupling capacitors and the physical contact based switching is achieved through micropoints or microflats. Both the realizations are achieved by deflecting a thin membrane either by employing thermal, piezoelectric or electrostatic actuation method.

Table 9.5 *RF switch requirements in cell phones (Source: Barbastathis et al.,* **Despande Center for Technological Innovation, MIT***)*

Technical requirement	*Cell phone key requirement*	*Available MEMS switch*
Voltage	3.3 Volts	30 (Too high)
RF Power handling	100 mW average, 4 W peak	2 W average, 8 W peak
Speed	20 mS	10 µS
Reliability	100 million cycles, 5 years	100 billion cycles, 3 years
Frequency	1–30 GHz	Up to 100 GHz

Thermal actuation: Thermal actuation uses Shape Memory Alloy (SMA). At low temperatures the crystalline structure of the alloy provides deformations. SMAs are active materials, which have the ability to return to a predetermined shape when heated. The SMA, when used for the design of actuator, is operated above the transformation temperature, a temperature below which it is susceptible to deformation. When the material is heated above the transformation temperature it undergoes a change in crystal structure, which causes it to return to its original shape. This phenomenon provides a unique mechanism for actuation. The commonly used SMA is Nitinol. Nitinol has very good electrical and mechanical properties, long fatigue life, and high corrosion resistance.

Piezoelectric actuation: Piezoelectric materials develop strain in an electric field. An applied voltage, known as bias voltage, across the piezoelectric material, generates the electric field. The elongation causes actuation. In effect the mechanical elongation of the material takes place. The piezoelectric material lead zirconate titanate (PZT) proved to be a viable material for high-speed switches. Piezoelectric actuation requires lower actuation voltages.

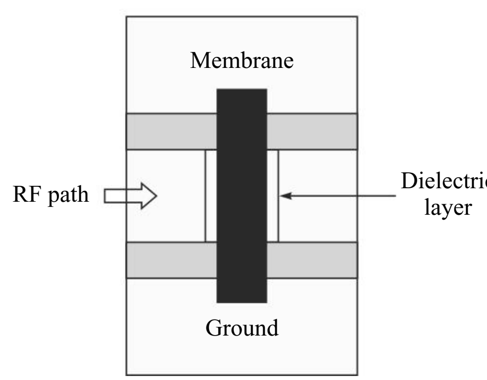


Fig. 9.16 *An RF MEMS switch*

Electrostatic actuation: In case of electrostatic actuation, the switch consists of a fixed-fixed thin metallic membrane suspended over a dielectric film that is deposited on the top of the bottom electrode. If an electrostatic voltage (called biasing or control voltage) is applied between the membrane and the bottom electrode, an electrostatic force is created to pull the membrane down. Electrostatically actuated RF MEMS switches require high actuation voltage, typically 30 volts or higher. Electrostatic actuation is the most used actuation scheme.

9.9.2 Capacitive-coupled Switch

Capacitive coupling switching is better than the contact based switching. The capacitive couplings are of two types; shunt coupling and series coupling as shown in Fig. 9.17. The signal flow from the previous stage (input port) to the next stage (output port) occurs only when the capacitance is optimized to match the impedance of both the stages. The matched capacitance is defined as the coupling capacitance (C_c). When no biasing or controlled signal applied to the electrodes of the capacitor the

value of the capacitance is away from the coupling capacitance. Properly deflecting the membrane obtains the desired value of coupling capacitance. Figure 9.18 shows schematic diagram of capacitive coupling based switch. Arrays of capacitors can be fabricated to obtain a matrix switch.

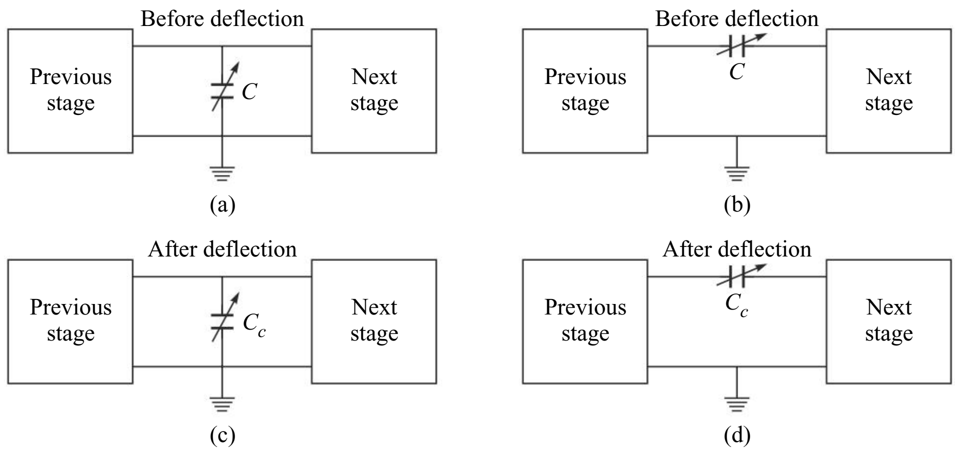


Fig. 9.17 *Capacitive coupling; (a) Shunt capacitive coupling (before the application of controlled voltage); (b) Series capacitive coupling (before the application of controlled voltage); (c) Shunt capacitive coupling (after the application of controlled voltage – now the coupling capacitance is C_c); (d) Series capacitive coupling (after the application of controlled voltage)*

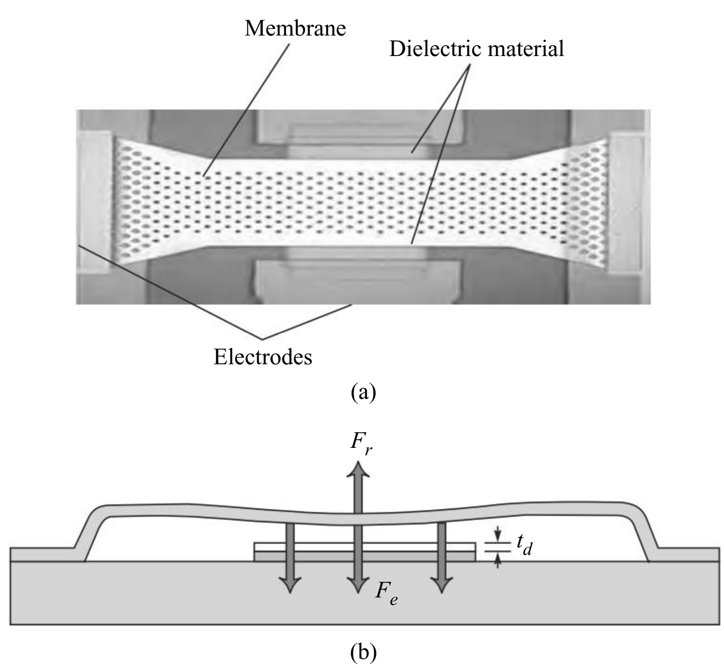


Fig. 9.18 *(a) Capacitive coupling based switch (top view), (b) Schematic of capacitive coupling based switch showing active forces and deflection*

Thickness of dielectric layer in capacitive coupling: The dielectric layer plays a key role as far as electrical isolation is concerned. Silicon nitride may be chosen as the dielectric material. Typical thickness is 0.2 µm (Fig. 9.16). The electrical isolation of the switch mainly depends on the capacitive coupling. For better isolation, the thickness of the dielectric layer should be smaller and at the same time the surface be smoother. However, the thinner dielectric layer causes the electric field intensity to increase significantly. Therefore, the thickness of the dielectric layer is chosen in such a way as to ensure that the electric field can never exceed the breakdown electric field of the dielectric material.

Fabrication: The fabrication steps are shown in Fig. 9.19. The design of a capacitive coupling based RF switch involves several different fields of knowledge, including mechanical design, RF design, and material science. The switches are fabricated by surface micromachining technique with a total of four masking levels. The micromachining processes involve five steps such as Ti/Cu seed layer deposition, dielectric material deposition, copper electroplating, aluminium deposition and release. The steps are briefly described as follows.

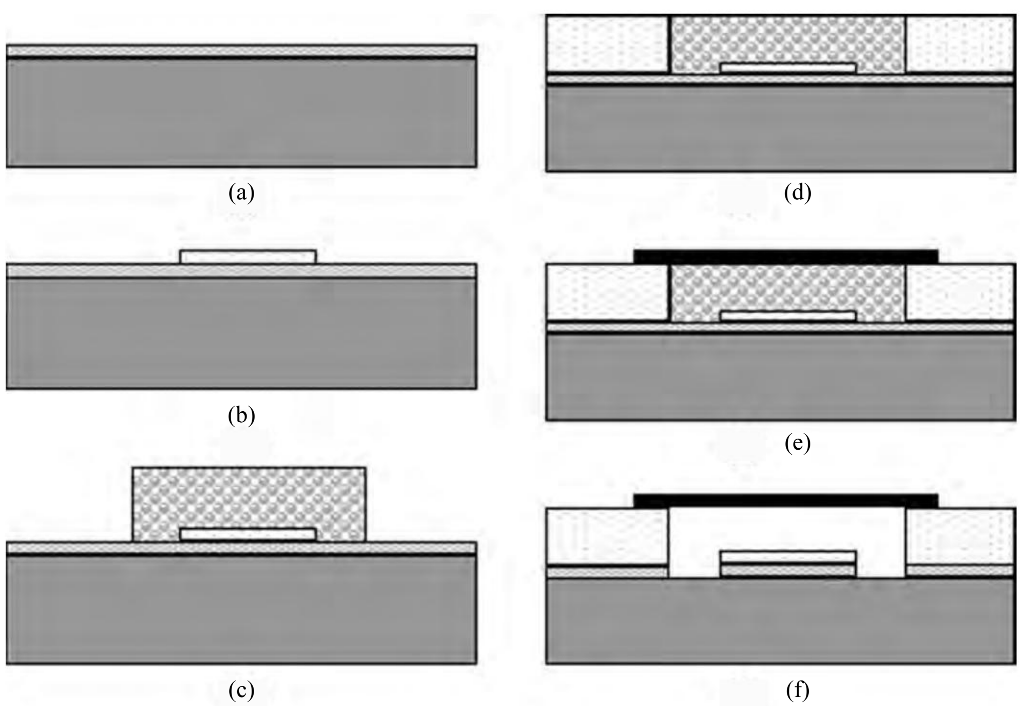


Fig. 9.19 *(a) Seed layer deposition, (b) dielectric layer, (c) Spacer coating and patterning, (d) Transmission line electroplating, (e) Membrane deposition, (f) Releasing phase*

- *Ti/Cu seed layer deposition:* The substrate can be a glass wafer. First a layer of titanium approximately 0.5 µm and then copper with thickness approximately 0.15 µm is sputtered on the substrate. This layer is called seed layer and is for electroplating.
- *Silicon nitride deposition:* A layer of dielectric material, i.e. silicon nitride of thickness 0.2 µm is then deposited and patterned. When in operation the dielectric material helps in blocking the DC

components. The dielectric material is deposited by the use of PECVD (Plasma-Enhanced Chemical Vapor Deposition) and Reactive Ion Etching (RIE) techniques.

- *Copper electroplating:* A photoresist layer is spin coated and patterned to define the electroplating area for the electrodes. At this stage approximately 4 μm thick copper layer is electroplated to define the strands (or posts) for the membrane.
- *Aluminium deposition:* This is the final deposition. A layer of aluminum of 0.4 μm is deposited by using electron beam evaporation and patterned. This layer determines the top electrode in the capacitor structure.
- *Release:* The photoresist sacrificial layer is removed to obtain the RF MEMS switch.

9.9.3 Contact Switch

Figure 9.20 shows a contact-type RF switch. There are two capacitors. They are called drive capacitors. The approximate dimensions of the capacitor plates are 100 by 100 μm. The thickness of the dielectric is about 2 μm.

The gap between the top and bottom plates is approximately 3.5 μm. The top membrane of the respective capacitors is suspended through spring-anchor mechanism called mechanical flexure. In between the two capacitors a microflat is present. The microflat is attached with the capacitor membranes. The microflat is called contact bridge. Below the contact bridge the RF line is present. When the contact bridge contacts RF line the RF circuit is closed and signal flows from one port to the other. That is the contact bridge makes it possible to achieve continuity across the signal line gap providing a low-loss transmission path for the RF signal. At off state the bridge device provides high electrical isolation between the input and output ports. The switch can be operated by applying voltage across the top and bottom drive capacitor plates, producing the attractive electrostatic force necessary to overcome the flexure spring force and to bring the contact bridge in contact with the underlying RF line. When the drive voltage is removed, the elastic energy in the mechanical flexures opens the switch. The pull-down force is approximately 60–80 μN.

Simulation tools are available to design and predict the performance of MEMS-based RF switches. ElecNet can be used to analyze 3D capacitive switches. ElecNet is a software tool package from Infolytica. Once the geometry and materials of the switch are entered, and voltages are applied to conducting parts, ElecNet calculates the distribution of electric potential around the conductors.

9.10 PHASE SHIFTER

A phase shifter is a two-port device, which alters the phase of its output signal in response to an input signal. Phase shifters have many applications in communication systems. A phase shifter changes the output signal phase by applying a variable control signal, usually a voltage. There are two types of phase shifters: constant and variable types. The constant type shifts the phase of the output signal to a value that is fixed. On the other hand a variable phase shifters can shift the phase of the output signal that is not fixed. Depending upon the demand the shifter can shift the output phase. The latter type are flexible in many respects. Further, there are two broad categories of variable phase shifters, namely analog and digital phase shifters. Analog phase shifters change the output phase by means of an analog control signal while a digital phase shifter uses a digital control signal.

One typical application of phase shifter is found in the SSB (Single Side Band) modulation and demodulation process. In SSB modulation scheme a phase shift of 90 degrees between the carrier

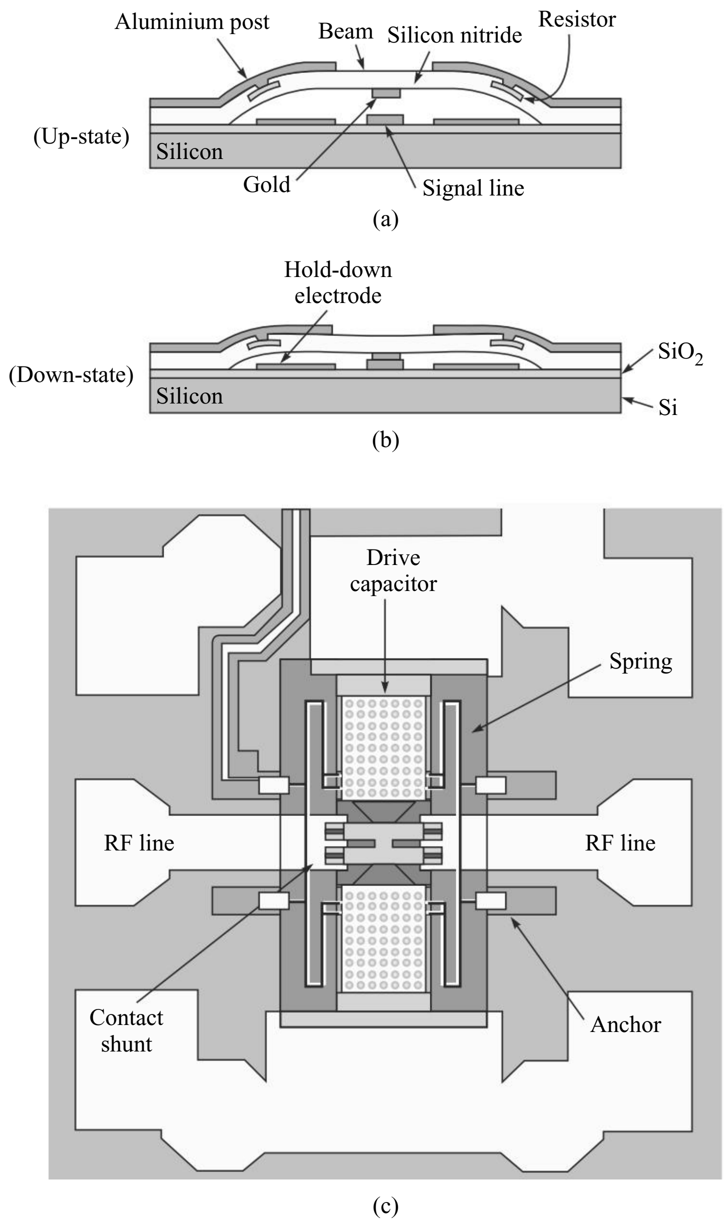


Fig. 9.20 *(a) Contact type switch before the application of control voltage, (b) After the application of voltage (Source:* Robeiz, Wiley, 2003*), (c) Top view*

signals is required in order to suppress one sideband out of two. Other important applications are found in satellite communication. Several satellite and terrestrial communication antennas require phased array antennas with beam steering and beam shaping capabilities. These phased array antenna systems use many phase shifters and RF switches. The role of phase shifter is to provide constructive or destructive interference so as to steer the radiated beam in the desired direction (which is described latter). There

are also other applications, in which 180-degree wideband phase shifting is used to flip an antenna or several antennas 180 degrees out of phase in relation to an array of antennas or vertical stack of antennas. This is useful in producing high take-off angles to match the propagation angles of a desired high frequency path.

9.10.1 Constant Type Phase Shifter

Fundamentally, phase shifters adjust transmission phase in a system. The phase shifting is achieved by changing the electrical path length as shown in Fig. 9.21. Line length 2 has a longer electrical length than line length 1. So when it is switched, the shifter will cause an in-line phase shift given by,

$$\Delta\Phi = \beta(l_2 - l_1) \tag{9.12}$$

where,

$$\beta = \frac{\omega}{V_p} \text{ and } V_p = \frac{1}{\sqrt{\varepsilon_{\text{eff}}}} \tag{9.13}$$

So we can design the micro-strip lines (phase shifter) such that line length 2 is longer than line length 1.

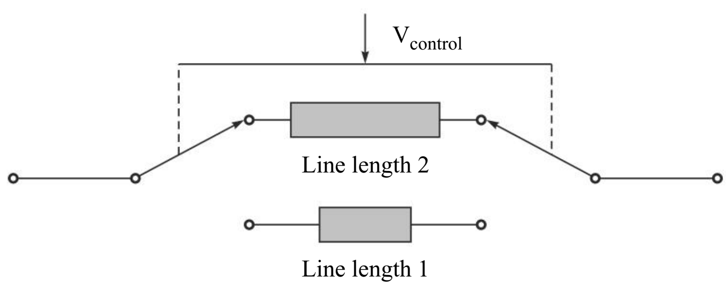


Fig. 9.21 *A switched line phase shifter*

Example 9.1 Design a switched-line phase shifter with a 22.5 degree phase shift at 4GHz, on a substrate with a dielectric constant of 9.9.

$$\Delta\ell = \frac{\Delta\phi}{\beta},\ 22.5 \text{ degrees} = \frac{22.5}{360}\cdot 2\omega = 0.392 \text{ rad}$$

$$@\ 4\text{GHz},\ \lambda_{\text{air}} = \frac{c}{f} = \frac{3\times10^8}{4\times10^9} = 0.075 \text{ m}$$

$$\lambda_g = \frac{\lambda_{\text{air}}}{\sqrt{\varepsilon_{\text{eff}}}} = \frac{0.075}{\sqrt{9.9}} = 0.0238 \text{ m} \qquad \beta = \frac{2\pi}{\lambda_g}$$

$$\therefore \qquad \Delta\ell = \frac{\Delta\phi}{\beta} = \frac{\lambda_g\cdot\Delta\phi}{2\pi} = \frac{0.0238 * 0.392}{2\pi} = 1.485\times10^{-3}\text{m} = 1.485 \text{ mm}$$

(Source: www.rfic.co.uk)

There are many configurations of phase shifters, but most operate as shunt switches placed at quarter wavelength intervals on a transmission line and configured to switch either a capacitance or an

inductance onto the line. This changes the electrical length of the line and produces a corresponding phase shift. A bright example of application of phase shifter is given below.

9.10.2 An Application of Phase Shifting

A phased array antenna (PAA) is composed of lots of radiating elements each integrated with a phase shifter that is controlled by the controlling voltage, $V_{control}$. Figure 9.22(a)&(b) show only two antenna elements. In Fig. 9.22(a) both radiating elements are fed with the signals having same phase. Note the three lobes (shaded portion) within which the radiating signals exist. In other words, the lobe represents the radiation pattern of a particular antenna configuration in a typical communication system. The middle lobe is the main lobe. To know more about the radiation patterns, refer a book on antennas and propagation, however, it suffices to say at this point that the radiation pattern in a multi antenna system is expressed in terms of lobes. In Fig. 9.22(b), the signal in the upper radiating element is radiated after being phase shifted by 10 degrees compared to the lower radiating element. Because of the phase

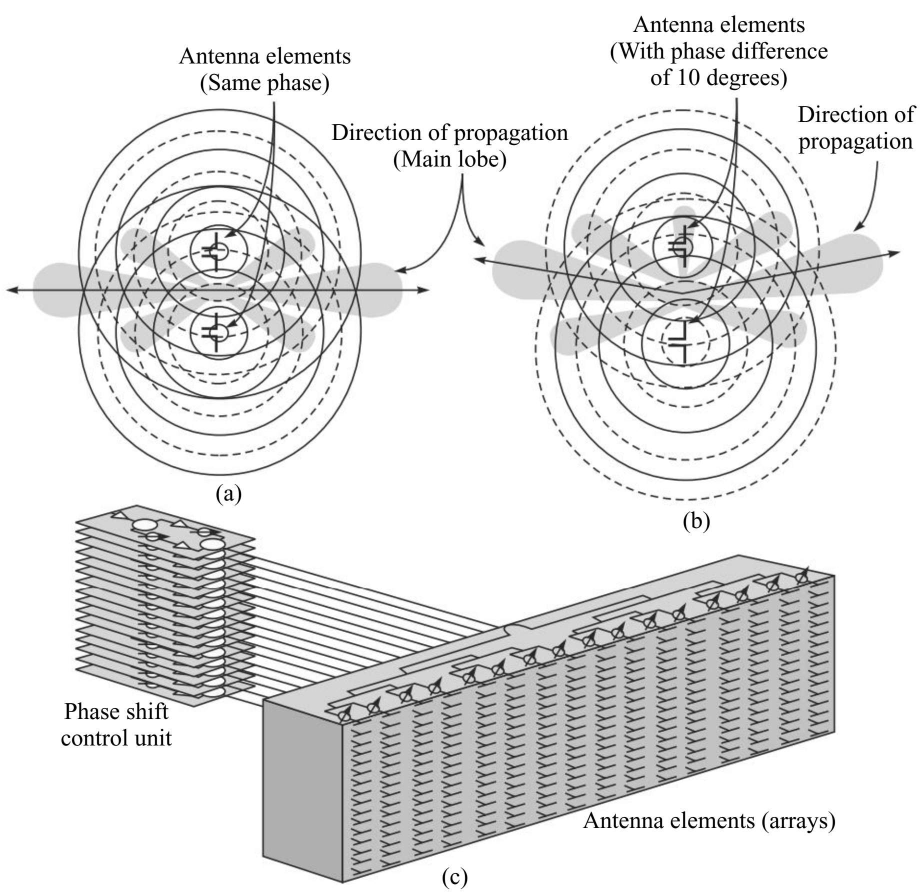


Fig. 9.22 *(a) Two antenna elements showing the direction of signal radiation (radiation lobe are shown in shaded portion) without phase shift; (b) A phase shift of 10 degrees have been applied. Now the direction of radiation changes; (c) A phased array antenna (PAA) elements whose phase can be controlled by control voltages so as to direct or receive the signal in a particular direction*

difference the direction of radiation will change (Fig. 9.22(b)). As you can see that the main lobe is a bit upward compared to that of Fig. 9.22(a). In principle, the signal is amplified by constructive interference in the main directions. Shifting the phase of the signal from the radiating element can amplify the beam in another direction due to the reason that the constructive interference now occurs in another direction.

9.10.3 Distributed Phase Shifter

In many phase shifters, the ultimate value of shifted phase is achieved sequentially. To achieve a phase shifting of 60 degrees ten phase shifters may be employed in series. These devices are called distributed phase shifters (DPS). Figure 9.23(a) shows the schematic diagram of a distributed phase shifter that contains ten bridges. Figure 9.23(b) is the photograph of the DPS and Fig. 9.23(c) is the cross-sectional view. The phase shifter is fabricated on a high resistivity silicon substrate using standard lithography.

The phase shifter consists of a high-impedance coplanar waveguide (CPW) transmission line. The bridge height is approximately 4.5 μm that can gives a phase shift of 6° at 18 GHz with a bias voltage of 30 V. The improved design of bridge can also provide a total of more than 90° at 18 GHz with similar bias voltages.

The CPW is a transmission line. A transmission line is an arrangement of two or more conductors or a waveguide that is used to transfer signal energy from one location to another. When the signal is fed into the CPW, the transmission line is said to be in loaded condition. The CPW has a central conductor. When a bias voltage between the bridges and the center conductor of the CPW is applied the height of

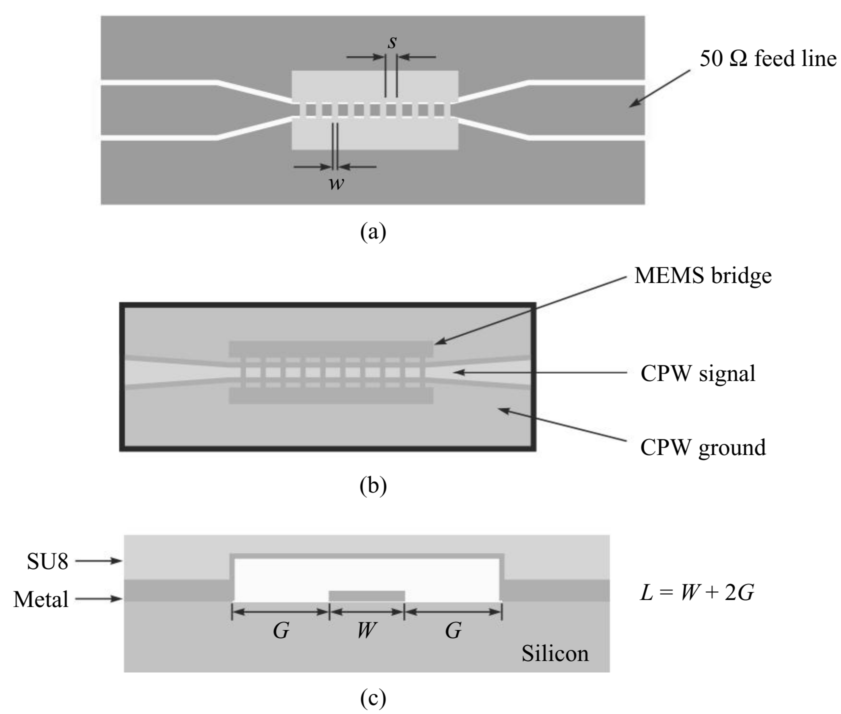


Fig. 9.23 *(a) Schematic diagram of a distributed phase shifter (DPS); (b) Photograph of the DPS; (c) Schematic illustration of cross-sectional view of the DPS (Source:* Ji et al., *Smart Mater Structure,* 10, Inst. of Physics, 2001*)*

the bridge is changed. This variation causes the change in the capacitance. The change in turn results in a change in characteristic impedance and phase velocity in this loaded transmission line. In effect, it provides a phase shift.

The applied bias voltage should be less than the pull-down voltage. Pull-down voltage is the maximum voltage at which the phase shifting property of the bridge collapses. So care must be taken in defining the dimension of the bridges vis-à-vis the bias voltage. Decreasing the bridge height or adopting bridge materials of relatively low elastic modulus such as polymer can reduce the actuation voltage. It is not recommended to reduce the height of the bridge as the parasitic capacitance may come into picture.

The maximum applied voltage that has to be kept less than the pull-down voltage is given by,

$$V_p = \sqrt{\frac{8k}{27\varepsilon_0 W\omega} g_0^3} \tag{9.14}$$

where ε_o is the permittivity of the free space, W is the width of the center conductor of the CPW, w is the width of the bridge, g_o is the height of the bridge and k is the effective spring constant of the bridge. The effective spring constant depends on the Young's modulus E and Poisson's ratio ν of the bridge material, the thickness t, the length L, residual tensile stress σ of the bridge and mathematically is expressed as,

$$k = \frac{32Et^3\omega}{L^3} + \frac{8\sigma(1-\upsilon)t\omega}{L} \tag{9.15}$$

The phase shift per unit length due to the change of the characteristic impedance arising from the bridges capacitance variation by applying a bias voltage is given by,

$$\Delta\Phi = 360 \times 10^{-3} \times \frac{fZ_0\sqrt{\varepsilon_{r,\text{eff}}}}{c}\left(\frac{1}{Z_{hu}} - \frac{1}{Z_{ld}}\right) \text{degree mm}^{-1} \tag{9.16}$$

where f is the frequency, Z_o is parameter that depends on the ratio of total width to the central conductor width, $\varepsilon_{r.\text{eff}}$ is the effective dielectric constant of the unloaded CPW transmission line and c is the free-space velocity, Z_{lu} and Z_{ld} are the characteristic impedance due to bridge capacitance at 0 V and the bias voltage, respectively.

A practical phase shifter comprises a 900 μm wide (W + 2G) CPW transmission line with 200 μm center (W) conductor, loaded periodically with ten MEMS capacitors (bridge) at a spacing of 590 μm and a height of 4.44 μm can provide distributed capacitance value of 98 fF mm^{-1} at 0 V and maximum capacitance of 150 fF mm^{-1} at the pull-down voltage of approximately 55 V, neglecting a fringe capacitance. The high-impedance line is designed to have a characteristic impedance of 76 ohm so that the loaded line has characteristic impedance close to 50 ohm.

9.11 SUMMARY

In the last two years, the development and deployment of devices based on microelectromechanical systems for radio-frequency applications have gained rapid ground. Microelectromechanical systems (MEMS) refer to a collection of microelements in terms of sensors and actuators, which can react to environmental change under microenergy control. The integration of MEMS for Radio Frequency (RF) applications has resulted in systems with superior performance levels and lower manufacturing costs. The incorporation of MEMS technologies into micro and millimeter wave communication systems

offers viable routes to MEMS actuators, antennas, switches and transmission lines. Such systems are called RF MEMS. The new technology have made discrete microscale mechanical circuits possible, which are capable of low loss filtering, mixing, switching, and frequency generation. Compared to conventional RF devices and products, RF MEMS operate with an increased bandwidth and increased radiation efficiency and have considerable scope for implementation within the expanding area of wireless personal communication devices including mobile phones.

This chapter therefore presents topical subjects covering the fundamentals, principles and applications of RF MEMS devices. In particular, the chapter starts with reviewing the communication systems. The important modules required for RF communication are then described. The advantage and some of the important design scenarios are explained. Comparisons of RF MEMS devices with respect to traditional RF devices are presented in a tabular form. The fabrication method for various types of inductors, the fundamental element of RF MEMS modules, is suggested. The usefulness, application areas, fabrication methods of varactors is covered. Further the principle of operation of resonators and their classifications are presented. RF MEMS switches are important building blocks in handing the transmission of RF power. The capacitive coupling and contact type switching are considered as the important RF switching methods. The operational methods and design criteria are described appropriately. Phase shifting is another functional attribute in the RF communication domain. Line phase shifter and distributed phase shifters are mostly used. In the final part of the chapter an example with respect to the need of phase shifting is presented.

Points to Remember

- RF stands for radio frequency.
- RF MEMS technology is a high-quality, three-dimensional, micro-scale structure that is manufactured using micromachining batch fabrication techniques.
- The important RF MEMS applications are found in (i) RF communication (Frequency of operation is tens to hundreds of kHz), (ii) Microwave communication (Frequency of operation is hundreds of MHz), (iii) Global positioning systems, and (iv) Broadband wireless access and wireless data link.
- Radio frequency microelectromechanical systems (RF MEMS) is a field that concerns with development of micromachined devices such as filters, oscillators, resonators, and switches, aimed at high frequency (~1 MHz to 60 GHz) communication applications.
- A communication system consists of three main building blocks such as transmitter, receiver and communication media.
- *Modulation* is a process of changing the parameters of a high frequency signal called *carrier signal*, with respect to the intensity of a given weak signal called *original baseband* signal or *modulating* signal.
- The basis of communication is to deal with the transmission of spectral power of the desired frequency component or band (group of consecutive frequency components) from one point to another in an effective way.
- Tuner is a frequency selective module that allows a band of frequencies to pass through.
- The transformation of maximum power will occur only when the input impedance of the tuner circuit equals the impedance of the air channel.
- Resonator, being a hollow chamber. The dimension of the hollow chamber plays important role in making resonance.

- A phase shifter is also a two-port device that alters the phase of its output signal in response to an external control signal. The phase shifting device is used to virtually flip an antenna or arrays of antennas in relation to producing propagation angles of a desired HF (high frequency) path.
- The purpose-made conductor coils are called inductors. Inductors are the basic building blocks of all types of oscillator, delay, and actuating circuits.
- The behavior of inductor in response to an alternating (ac) signal is that it opposes the input current change.
- When electromagnetic RF signal flows through RF inductor an electromagnetic field is established and it changes in accordance with the input signal. The changing electromagnetic field causes an induced voltage in a direction opposite to the flow of input signal.
- The quality Q, of an inductor is the measure of inductiveness of the coil. In other words, Q expresses the merit of the inductor.
- Some RF applications require current to be confined to the surface of the conductor. One such building block is the antenna.
- Varactor is an active device whose capacitance value is varied by some means. Standard MEMS actuation principles such as electrostatic or thermal actuation can be employed to vary the capacitance.
- The relative merits of MEMS varactors are low power consumption, high quality factor, tolerance for high voltage swing, low harmonic distortion and large tuning range.
- A simple capacitor-inductor circuit can act as the tuner for the RF receiver.
- LC tank circuit can generate a single frequency signal. However, resonators are popular where the need for high stability and selectivity arises.
- A switch is a device consisting of a mechanical, electrical, electronic or optical element for making or breaking or changing the connections in a circuit.
- Waveguide and coaxial switches offer the benefits of low insertion loss, good isolation characteristics and high power handling capabilities. However, they are bulky, heavy, and slow. The RF MEMS switches significantly reduce the weight, size, power consumption and parts count.
- MEMS switch is based on either capacitive coupling or physical contacts. Both the realizations are achieved by deflecting a thin membrane either by employing thermal, piezoelectric or electrostatic actuation principle.
- A phase shifter is a two-port device, which alters the phase of its output signal in response to an input signal.
- There are two types of phase shifters: fixed and variable.
- There are many configurations of phase shifters, but most operate as shunt switches placed at quarter wavelength intervals on a transmission line and configured to switch either a capacitance or an inductance onto the line. This changes the electrical length of the line and produces a corresponding phase shift.
- In many phase shifters, the ultimate value of shifted phase is achieved sequentially. To achieve a phase shifting of 60 degrees ten phase shifters may be employed in series. These devices are called distributed phase shifters (DPS).

Exercises

1. What does RF stand for? Write down the frequency range of RF signals. Also mention the sub-ranges, their designations and corresponding wavelength ranges.
2. Draw the block diagram of a general communication system.
3. What do you mean by modulation and demodulation process? What are the types of modulation techniques usually employed in communication systems?

4. List out the important RF modules and briefly explain their functions.
5. Suggest the application domain of RF MEMS. What are the important advantages of using RF MEMS as compared to traditional units and systems?
6. Mention some of the design scenarios of RF MEMS you know.
7. What is an inductor? What types of MEMS inductors do you know? Draw the equivalent circuit of a typical planer inductor. Define Q of an inductor. Describe the important process sequences as far as fabrication of MEMS planer inductor is concerned.
8. What is the primary reason of designing solenoid type MEMS inductor? Write down the diminutions and specifications of a typical MEMS solenoid inductor.
9. What is a varactor? List the advantages of MEMS varactor over the traditional varactor. Explain the theory behind the varactor. Write down the expression governing the electrostatic force generated between the varactor plates by the applied tuning voltage and define the terms used therein.
10. Describe the important process sequences as far as fabrication of MEMS varactor is concerned.
11. Write notes on the following.
    (a) Tuner (b) Resonator
12. What do you mean by capacitive coupling? Discuss the principle of operation of a typical MEMS switch used for RF applications. What are the various types of actuation methods used in RF switching? Describe the important process sequences with regard to fabrication of RF MEMS switches. Differentiate coupling-switch from the contact-switch.
13. What is a phase shifter? Mention the applications of phase shifters. Explain the principle of operation of a switched-line phase shifter.
14. Give an exclusive application of a phase shifter (Hint: Phased Array Antenna).
15. What is the meaning of distributed phase shifting? How is this achieved? Explain with suitable diagram. Discuss the MEMS DPS.

Chapter 10

Microfluidic Systems

Objectives

The objective of this chapter is to study the following.

- Microfluidic systems (MFS)
- The concept of Lab-on-a-chip
- Important considerations on microscale fluids and systems
- Properties of fluids
- Various phenomenological effects for fluid flow. In particular, the following fluid flow phenomenon and their principles of operations are presented.
  - Dielectrophoresis (DEP)
  - Electrowetting
  - Electrothermal
  - Thermocapillary
  - Electroosmosis
  - Optoelectrowetting (OEW): Light-actuated fluid flow
- MEMS channel, filter and micropump

10.1 INTRODUCTION

The study of motion or transportation of fluids and their mixtures at a microscale level is known as microfluidics. Microdevices, which are used to transport and store fluid are called microfluidic systems (MFS) or microfluidic devices (MFD). Typically, the MFS handle fluid volumes in the order of nanoliter. Microfluidic systems are the subject of great scientific and commercial interest for a wide range of applications, including the biomedical, environmental, automotive, aerospace, and defense.

Microfluidic system is a recent development, which obviously refers to the MEMS devices capable of handling small volumes of fluids in the order of nano, pico and even femtoliter volumes. Microfluid devices are designed to inject, move, separate, and merge liquid droplets. Sometimes, the microfluidic devices have dimensions ranging from millimeter range to micrometers. Confusion arises in defining microscale device and microfluidic device, but in reality the latter one is much smaller than the former one, however, these terms are used synonymously.

Microfluidic devices involve construction and design that differ from macroscale fluidic devices. Many conventional macroscale devices have already been in use however; microfabrication conformant devices and systems can provide greater performance over conventional devices. Nevertheless, the microfluidic devices are fabricated by micromachining techniques that fulfil the very stringent requirements on reliability and compatibility issues. Design and development of microfluidic devices entails basic knowledge on the properties and dynamics of the fluids. The study also demands to acquire knowledge as far as interaction of fluid with the material and with the electrical, thermal and optical signals are concerned.

This chapter considers the fundamentals at the core of macroscale fluid mechanics. Important criteria such as validity theory of macroscale fluid dynamics will be presented. This includes the study of continuum phenomena. Further, the importance of surface tension in both continuous and discrete flows will be the matter of discussion. Broadly the following topics will be presented.

- Concept of lab-on-a-chip
- Properties of fluid such as density, viscosity and surface tension
- Surface wetting phenomenon such as dielectrowetting (DEW), thermalwetting, thermocapillary wetting, electroosmosis flow (EOF) and optoelectrowetting (EOW)
- Microchannel, molecular gate as filter and micropump

10.2 APPLICATIONS

Microfluidics leads to establish a foundation knowledge on modeling, developing and designing typical microfluidic devices such as the microcapillary tubes, channels and systems which are eventually used in many fields like life sciences, chemistry, medicine, computing (inkjet printing) to name a few. These devices offer many advantages, including small sample volume requirement for expensive or hazardous reagents and short diagnosis time. It is considered as a very important enabling technology in the field of drug development and testing where delivery of minute amounts of liquid to distinct reaction vessels and substrates is required. Some of the specific applications of microfluidics are,

- Ejection of inkjet droplets in printers
- Microfluidic oscillator and micro heat exchangers
- Tuning of optical-fiber properties
- Micropumping of gases and liquids
- Drug screening and delivery
- In-vitro diagnostics
- Biological and genetic analysis (e.g. DNA detection)
- Chemical analysis and synthesis
- Environmental pollutant detection and analysis

10.2.1 Lab-on-a-chip

In most of the chemical, clinical and bio-clinical laboratories the instrumentation and analytical equipment and devices are becoming smaller, simpler, and smarter suggesting miniaturization. The micro handling of fluids in such laboratories is a significant development. Microfluidics implicitly concerns the design and development of microscale devices for medical, chemical and biotechnological research. These devices are rather called tools, which have been emerged as biochip or lab-on-a-chip. This nomenclature is from the fact that the advanced miniaturized technology replaces the traditional

macroscale test-tube based instrumentation and analytical equipment wthin the laboratory. The lab-on-a-chip is considered as a small laboratory, which allows the scientists to perform experiments in a small place. The advantages are,

- Smaller liquid consumption
- Good response time
- Faster analysis and diagnosis
- Better statistical results and certainty
- Improved possibilities for automation
- Decrease in health and environmental risks
- Reduced costs

10.3 IMPORTANT CONSIDERATIONS ON MICROSCALE FLUID

The behavior of fluid is significantly changed as geometric scale decreases. In this respect following considerations are to be noted.

- The physical, technological, and biological significance in flows of gases and liquids at the microscale level necessitate the study of properties of fluids.
- The dominant physical quantities change in the micro-world.
- Because of scaling effect the large surface forces, high shear and extensional rates, high heat and mass transfer rates make microfluidics a challenging technology.
- The control of fluid flow in miniaturized devices and porous media is critical.
- The physical phenomenon such as intermolecular forces, slip, diffusion and bubbles are the main active agents at the microscale level.
- In the microworld the surface forces and surface tension start to dominate. When the channel of the order of one micrometer the surface tension is extremely large.
- The design of MFS concerns selection of appropriate method for inletting or pumping the liquids into microchannels against two major forces such as surface tension and the externally applied pressure.
- The diffusion-based characteristics of the laminar flow are sometimes exploited for sample preparation and analysis.
- The laminar flow behavior of fluids is also considered in the design and development process of microfluidic devices.
- The Newtonian fluid mechanics and flow in confined geometries are significant.
- The flow of thin films spreading under gravity or surface tension gradients is considered important.
- The handling of fluids with liquid-gas interface in micro channels, valves, pumps, mixers, separators and reactors, excels engineering and scientific challenges.
- Fluid transportation in a typical microchannel is accomplished based on many phenomenological methods (described latter).
- The control of fluid transportation apparently depends upon the wall surface physicochemistry due to the fact that the fluids exert hydrophobic or hydrophilic force from the channels.
- The flow behavior of fluids is influenced by the presence of ions, polymers, biomolecules, etc.

10.4 PROPERTIES OF FLUIDS

Analysis of fluid motion is performed in the same way as in solid mechanics, but it draws special attention to understand the properties of fluids when in motion. The properties and essential parameters involved in fluid mechanics are briefly reviewed below, but the readers are advised to follow a standard book for further topical knowledge. Prior to review we prefer to define the important laws and terms those are frequently used in this domain.

- *Bubbles:* The surface tension provides the required tension in order to form a bubble. The bubbles are formed to satisfy the requirement of minimum energy. In other words, the tendency to minimize wall tension brings out the bubbles into a globular shape.
- *Droplets:* Here also surface tension plays a role for shaping of liquid droplets. The droplets are easily deformed. One interesting viewpoint is that a droplet tends to form a globular shape by the cohesive forces of the surface layer.
- *LaPlace's Law:* Wall tension is developed to confine a fluid of given pressure. The wall tension increases with the radius of the container for a given internal fluid pressure. A spherical container has half the wall tension of a cylindrical type, for a given radius and internal pressure.
- *Static fluid pressure:* The static fluid pressure depends upon the height, density, and gravitational force.
- *Energy density:* Pressure in a fluid is a measure of energy per unit volume called energy density.
- *Undiminished pressure:* Pressure is transmitted undiminished in an enclosed static fluid. This is called Pascal's principle.

10.4.1 Density

Let us consider a mass that is contained in a finite volume of fluid. We now can define the average mass density of the fluid as,

$$\rho = \lim_{\delta v \to \delta v'} \frac{\delta m}{\delta v} \tag{10.1}$$

where, ρ is called density, δm and δv are the mass and volume. If the fluid is considered as composed of discrete particles, the instantaneous density at a point in the fluid can take different concept due to the fact that the mass is now not a continuous function of volume and therefore, the limit of the above ratio as volume approaches zero does not exist. This situation in turn implies that the mass should be viewed from another approach, which suggests that mass is regarded as the actual mass contained in the volume at an instant but as the probable mass which are expected to find in the volume at that instant. The most probable value depends upon the statistical assemblage of molecules making up the fluid. Under this approach, it can then be presumed that there exists a density function for any arbitrary volume of fluid such that

$$m = \int_v \rho\,(x, y, z, t)\, dV \tag{10.2}$$

The evaluation of this function is difficult, for which the real fluid is always considered as a continuum fluid. In essence, density is the mean value for a finite volume and can be determined.

10.4.2 Viscosity

The fluid is defined as a substance that cannot support shear force in steady state as a result shear stress is developed. A fluid continuously deforms or flows under shear stress. The viscosity of a fluid determines its ability to support shear forces.

10.4.3 Nature of Flow

Fluid flow is characterized by laminar flow and turbulent flow. The laminar flows are also called layered flow as the fluid stream flows parallel to each other. The Reynolds number quantifies the two types of flow and is defined as the ratio of inertial to viscous force of the fluid flowing in the channel. This is also equivalent to the ratio of momentum to the frictional force exerted on the fluid by the channel surface. Mathematically it is written as,

$$\Re e = \frac{\rho v d_c}{\mu} \tag{10.3}$$

where, ρ is density, v is the velocity, d_c = characteristic depth of flow and μ is viscosity. The effective diameter depends upon the geometry of the channel. The laminar flow is characterized by low Reynolds number signifying the flow is dominated by viscous forces, while the Reynolds number is high in case of turbulent flow and the typical value is more than 2000. If the boundary conditions are kept constant, then the fluid velocity at all locations are constant. In turbulent flow the inertial forces dominate and the flow is random in both space and time.

10.4.4 Surface Tension

The surface tension is of great importance at microscale dimension. The liquid under investigation could be a microscopic liquid film, which affects interaction with the surface. The molecules, which make up a liquid, attract each other. An equal attractive force in all directions balances the interactions of an individual molecule. Molecules on the surface, however, experience an imbalance of forces. The liquid molecules arrange themselves in such a way that the cohesive forces between them are aptly shared by the neighboring molecules, leading to zero net force. The situation is completely different if the liquid comes in contact with the surfaces. In fact the molecules do not find scope to arrange themselves, as usually they do, due to the presence of neighboring surface and therefore the normal arrangement gets disturbed leading to non-zero net force. The effect is balanced by a phenomenon what is known as the surface tension. The non-zero net force thus developed can be calculated by dividing the length along the liquid surface. In other words, surface tension is a measurement of the cohesive energy present at an interface of two surfaces (another surface could be air). The surface tension is expressed in force per unit width, which is dynes/centimeter (dynes/cm.) or milliNewtons/meter (mN/m). Surface tension of some liquids against air or vacuum exposed to vapor pressure of the substance is provided in Appendix B.

Since the molecules on the surface of a liquid experience an imbalance of forces because of missing bonds of some sort, the net effect of this situation is the storage of free energy at the surface of the liquid. Surface tension tries to minimize the surface area, forming spherical droplet. At the interfacing, the free energy is always proportional to its surface, which can be expressed as,

$$E = \sigma S \tag{10.4}$$

where, E is the free energy, S is area of the surface, and σ is called coefficient of surface tension (dyne/cm). It is constant as long as the surface force has no tangential component. Figure 10.1 shows a look of a microvolume of fluid when placed on a solid surface. The angle between the tangent to the curvature and the solid surface, what is known as contact angle, can be obtained from the following formula. This is known as Neumann formula. The Neumann formula is applied when one of the surfaces is solid.

$$\sigma_2 - \sigma_1 = \sigma_3 \cos\theta \tag{10.5}$$

where σ_1, σ_2 and σ_3 are shown in Fig. 10.1. θ is called contact angle.

In order to minimize the free energy the curvature will tend to decrease. Assuming the surface tension coefficient constant (i.e. surface force has no tangential force), then the difference between inside and outside pressure of a liquid is expressed as

$$\Delta P = \sigma\kappa \tag{10.6}$$

where κ is called mean-free surface curvature. The above formula satisfies the Laplace law and is useful for studying the static droplet with surface tension acting along its interface with no gravity force. The curvature of the static droplet is calculated as $\kappa = 1/r$, where r is surface curvature radii in orthogonal planes.

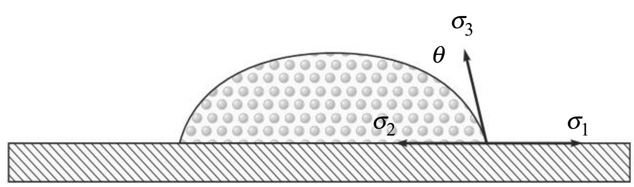


Fig. 10.1 *Contact angle in case when one of surfaces is solid*

The effect of microfluidic surface tension can be utilized in many applications. The effect can be employed to act as the medium of power transmission, as in conventional fluid power systems. Surface tension can be utilized in many microactuating systems. Surface tension is helpful in designing microbearing systems.

10.5 ANALYTICAL EXPRESSIONS FOR LIQUID FLOW IN A CHANNEL

The analytical expressions for propagation of the liquid in a channel are provided in this section. The incompressible, quasi-steady, laminar and Newtonian $2D$ horizontal flow of liquid is expressed what is known as Navier-Stokes equation. In fluid dynamics the Navier-Stokes equations are a set of nonlinear partial differential equations that describe the flow of fluids.

$$-\frac{\partial P}{\partial x} = -\mu \frac{\partial^2 V_x}{\partial y^2} \tag{10.7}$$

where P is the pressure of the fluid at x as mentioned in Fig. 10.2, μ is the viscosity of the fluid in the channel, V_x is the velocity. The coordinate x and y are oriented as shown in the figure. The equation reflects the conservation of momentum. Each side of the above equation is independent of the other. Therefore, the partial differential equation can be separated into two different equations.

$$\frac{\partial P}{\partial x} = -\beta \quad \text{and} \quad \mu \frac{\partial^2 V_x}{\partial y^2} = -\beta \tag{10.8}$$

where β is a constant. By applying appropriate boundary conditions, the advancement of the liquid front in the channel can be expressed in terms of separation distance, flow time, surface tension, contact angle, and the viscosity.

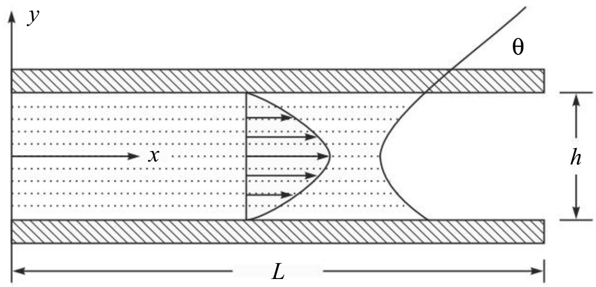


Fig. 10.2 *Flow between parallel plates*

$$L = \left(\frac{ht\sigma\cos\theta}{3\mu}\right)^{\frac{1}{2}} \Rightarrow t = \frac{3\mu L^2}{h\sigma\cos\theta} \tag{10.9}$$

Equation 10.9 suggests that the flow time t, is inversely proportional to the surface tension σ, separation distance h, and the cosine of the wetting or contact angle θ and directly proportional to the viscosity μ and square of the flow distance L. The above solution is obtained by ignoring the effect of surface roughness, flux residues, and flow obstructions and by assuming an infinite supply of liquid without any end effect.

For compressible and viscous fluid, a set of the equations can be written as follows:

$$\rho\frac{Du}{Dt} = \rho f_x - \frac{\partial p}{\partial x} + \mu\nabla^2 u \tag{10.10a}$$

$$\rho\frac{Dv}{Dt} = \rho f_y - \frac{\partial p}{\partial y} + \mu\nabla^2 v \tag{10.10b}$$

$$\rho\frac{Dw}{Dt} = \rho f_z - \frac{\partial p}{\partial z} + \mu\nabla^2 w \tag{10.10c}$$

where, D/Dt is the *material derivative*, ρ is density, u, v, w are the three velocity components, f_x, f_y, f_z are the components of forces, μ is viscosity of a fluid and is uniform, ∇ is the del operator. The continuity equation is given by,

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) + \frac{\partial}{\partial z}(\rho w) + \frac{\partial \rho}{\partial t} = 0 \tag{10.11}$$

10.5.1 Computational Fluid Dynamics

An understanding of the motion of fluids is crucial. With the rapid growth in processing power, software applications now bring numerical analysis and solutions of flow problems. The solution to momentum and continuity equation through computer is referred to as Computational Fluid Dynamics (CFD). These equations are solved numerically yielding a complete picture of the flow. Apparently, CFD is a computational technology that enables to study the dynamics of things that flow. Undoubtedly, CFD is a sophisticated analysis technique. Using CFD, computational model that represents a fluid system can be built. Through the application of fluid theories the corresponding algorithms are developed which can provide the prediction information on the fluid dynamics. CFD-based analysis can predict flow patterns in order to optimize the design and operating conditions. Within the algorithm, partial differential

equations describing the fluid flow such as the Navier-Stokes equations are rewritten as algebraic equations that relate the physical parameters such as sample concentrations, pressure, velocity and temperature. Since CFD concerns the dynamics that flow its scope is extended to study transfer of heat, mass, stress analysis and chemical reaction.

10.6 FLUID ACTUATION METHODS

The microfluidic systems are of two types based on the way the microvolume fluid is transported (or its position is manipulated). Accordingly, the systems are called,

- Continuous flow systems
- Liquid droplet-based system

The position manipulation of microvolume liquid is sometimes called fluid actuation. Conventional pumps, valves, and channels actuate the fluids in continuous-flow systems. In a droplet-based system, however, they are actuated by exploiting the surface tension. In essence, the systems use surface tension gradient to move, combine, and mix liquid droplets. The latter system is also called *digital* fluidic microsystem.

10.6.1 Digital Microfluidics

The application of a potential difference between the sandwiched liquid and the electrodes changes the hydrophobic[1] or hydrophilic[2] character of the region. This, in turn, causes the transportation of the liquid along the region, where it can then be separated into a smaller segment. That part of the liquid can then be driven further by sequential potential application. This procedure results in a digitized fluidic circuit. The procedure has attracted much attention because it eliminates the need for traditional pumps and valves. The method minimizes the risks of cross contamination, and also further reduces volume of samples under investigation.

The development of fluidic microsystems requires fundamental study on actuation mechanisms. Surface tension is the dominant force in microscale and has been used widely in actuating microvolume fluid and controlling the liquid. There are several fundamental mechanisms by adopting which the fluid can be actuated. All these mechanism can control the surface tension. Some of the important mechanisms are:

- Dielectrophoresis
- Electrowetting
- Electrothermal
- Thermocapillary
- Electroosmosis
- Light-actuated microfluidic device called optoelectrowetting

The main advantage of these actuation methods is that they can also represent open channel system technology. Open systems facilitate easy fabrication, cleaning and experiment observation.

10.7 DIELECTROPHORESIS (DEP)

Pellat conducted the first experiment in connection with the observation of dielectrophoretic (DEP) phenomenon. He did this by utilizing two planar, parallel and opposed electrodes, placed vertically with one end submerged into an insulating, dielectric liquid as shown in Fig. 10.3(a). From the experiment it was found that if a potential difference between the two electrodes is applied, a force is exerted on the

[1] Immiscible with water
[2] Strong affinity for water

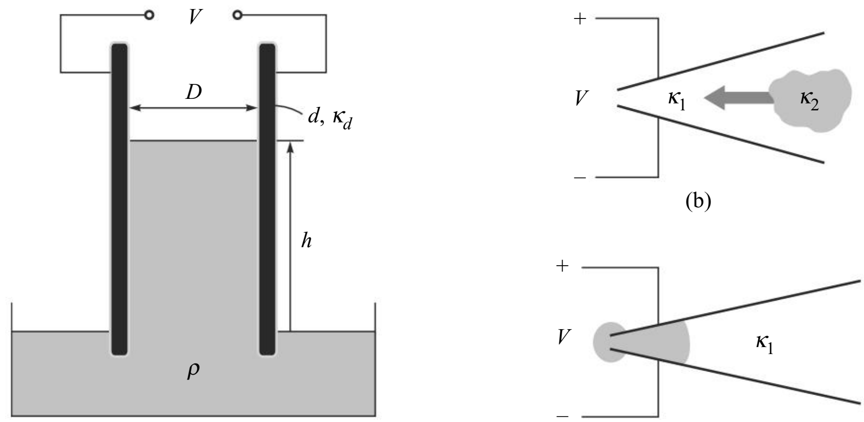


Fig. 10.3 *(a) Pellat's dielectrophoretic force experiment; (b) The dielectric liquid is attracted to regions with stronger electric field; (c) The liquid surface follows the electric field lines*

liquid, trying to impel it upward. As mentioned earlier, because of applied potential difference, the hydrophobic and hydrophilic character of the region changes. This causes the development of a force. The magnitude of this force, and therefore the height of rise of the liquid, is proportional to the magnitude of the applied voltage. The mathematical equation that governs the DEP phenomenon can be written as,

$$h \approx \frac{k_d \varepsilon_0 V^2}{4 g \rho d D} \tag{10.12}$$

where V is the applied potential difference, ρ is the density, k_d is the dielectric constant of the liquid, h is height of rise of liquid, d is the thickness of the dielectric coating, D is the distance between the planar-parallel electrode, $e_o = 8.854 \times 10^{-12}$ Farad/meter is the permittivity of the free space, $g = 9.81$ m/s^2 is the gravitational force. The effect will be similar if the electrodes are not placed parallel as shown in Fig. 10.3(b). The force that starts to act on the liquid is called dielectrophoretic force. The dielectrophoretic force appears when a medium (liquid) is exposed to a non-uniform electric field. The liquids are pulled toward the regions of stronger electric field (Fig. 10.3(c)). In essence, the dielectrophoretic phenomenon can be utilized in microfluidics to transport and manipulate microvolume of liquid on a surface (channel).

10.8 ELECTROWETTING

For manipulation of liquid on surface, electrowetting effect is sometimes preferred. Electrowetting is equivalent to the conventional pumping mechanism. It is a phenomenon whereby an electric field can modify the wetting behavior of a droplet that is in contact with an insulated electrode (Fig. 10.4). The principle behind the electrowetting method is that the contact angle of a partially wetting conductive liquid on an insulating substrate can be reduced by applying a voltage between the liquid and a counter electrode sitting underneath the insulating layer as shown in Fig. 10.4(b). Figure 10.4(b) shows the electrowetting setup with conductive liquid, insulating substrate and counter electrode. The contact angle decreases as electric field is applied. θ_1 is the contact angle prior to the application of the electric field and θ_2 is the contact angle after the application of the electric field. Note that $\theta_1 < \theta_2$.

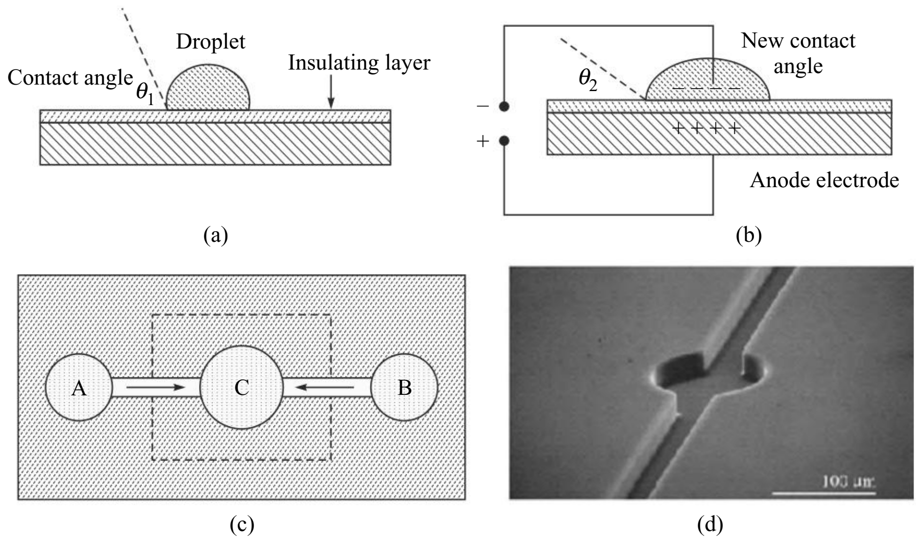


Fig. 10.4 *(a) Schematic of a microfluidic surface system consisting of a dielectric material and substrate, (b) Electrowetting setup, (c) Liquid is deposited in the reservoirs A and B and pumped along the hydrophilic channels into chamber C, where a chemical reaction can takes place, (d) SEM picture of a channel and well of the portion of the figure, (c) marked dotted*

The relationship between the contact angle and the applied voltage can be expressed as:

$$\cos[\theta(V_i)] = \cos[\theta(0)] + \frac{1}{2}\frac{\varepsilon}{t\gamma_{LV}}V_i^2 \tag{10.13}$$

where V_i is the applied voltage across the insulator, t is thickness of the insulating layer, ε is dielectric constant, and γ_{LV} is the interfacial tension between liquid and air. If AC signal is applied, V_i should be RMS value.

Electrowetting arrays allow large numbers of droplets to be independently manipulated under direct electrical control. By employing suitable pattern of electrodes, a very successful integrated channel can be designed in order to manipulate both shape and position of liquid. In electrowetting the type and nature of electrodes along with the value of applied voltage determines the efficient and proper operation of liquid manipulation, else instability may occur. For instance, if the applied voltage is too high the microvolume liquid may emit small droplets called satellite droplet. Another instability is called self-excited oscillations, which means that for certain applied voltage a capillary neck is formed that break and reform periodically. Electrowetting method can efficiently be utilized for fluid mixing applications (Fig. 10.4(c)).

The distinguishing feature of electrowetting and dielectrophoresis (DEP) is their ability to shape a liquid mass on a substrate. For the coplanar electrode geometry, the DEP force creates a semi-circular profile. In many other non-microfluidic conventional applications the principle of DEP has been applied for achieving the movement of particle around in a liquid by applying a voltage and getting them to move. On the other hand, electrowetting gets liquids to move around in other liquids (not the particle within the liquid). Electrophoresis is a slow effect. Electrowetting devices are quick, approximately the movement up to 50 cm/sec is observed although 10 cm/sec is more typical. Uniformly sized droplets of the order of 10 picoliters can be dispensed accurately and rapidly.

10.9 ELECTROTHERMAL FLOW

Electrothermal flow arises from the temperature gradient in the medium generated by joule heating of the fluid. Because of the gradient, parameters such as conductivity, permittivity, viscosity, and density of the liquid are changed, establishing gradients in these parameters within the liquid. The gradients then make it possible to generate forces, which eventually act on the fluid. For example, a permittivity gradient produces dielectric force while the conductivity gradient produces Coulombic force.

The governing formulae can be derived as follows. The heat is obtained from the electrical energy. The electrical field applied to manipulate microvolume fluid signifies that there is a power density associated with the electrode, thereby with the surrounding fluid. The power transferred per unit volume is given by,

$$W = \sigma E^2 \tag{10.14}$$

where σ is electrical conductivity, W is power in Wm^{-3}. The power is dissipated from the electric field and equivalent power is generated in the fluid surrounding the electrode. Because of the power generation, the temperature rises. By solving the power equation, the rise in temperature can be approximated to:

$$\Delta T \approx \frac{\sigma V_{rms}^2}{k} \tag{10.15}$$

where V_{rms} is the root mean square voltage, k is a constant called thermal conductivity. At the same time the analytical expression of the force generated because of the effect is estimated to be,

$$f_e = -M(\omega,T)\left(\frac{\varepsilon\sigma V_{rms}^4}{2\pi^3 kr^3 T}\right)\left(1-\frac{2\Phi}{\pi}\right) \tag{10.16}$$

where $M(w,T)$ is a dimensionless factor. This factor describes the variation of the electrothermal force as a function of applied frequency and given by the formula,

$$M(\omega,T) = \left(\frac{\frac{T}{\sigma}\frac{\partial\sigma}{\partial T} - \frac{T}{\varepsilon}\frac{\partial\varepsilon}{\partial T}}{1+(\omega\tau)^2} + \frac{1}{2}\frac{T}{\varepsilon}\frac{\partial\varepsilon}{\partial T}\right) \tag{10.17}$$

where w is the angular frequency, ε is the fluid permittivity, r and ϕ are the variables in polar coordinates system with reference to the middle of the electrode gap. Through suitable estimation, appropriate electrodes and microstructures can be designed to manipulate the microvolume fluid. Eventually, the manipulation is based on electrothermal flow.

10.10 THERMOCAPILLARY EFFECT

There is a similarity between the electrothermal and thermocapillary flow, although in the latter case the effect is on a capillary tube. The inkjet printing technology is based on the thermocapillary effect. The thermocapillary effect facilitates the formation and delivery of tiny drops at precise locations on the receiver material (e.g. paper). Heating a fluid meniscus (curved surface of a liquid at the liquid-air interface in a microtube) non-uniformly induces a gradient in surface tension. This gradient produces a tangential force called *Marangoni force*, on the free surface of the liquid. The force is also defined as

thermocapillary-driven force. Whenever fluid dimensions are less than 10 microns, thermocapillary-driven forces make it possible to separate the droplets discreetly from the main fluid body contained in the microcapillary tube. The droplets are propelled through space in front of which the printing materials (e.g. paper) are placed. At the microlevel the influence due to gravity is extremely small for which the discrete droplets, while propelled through the space are not disturbed.

10.11 ELECTROOSMOSIS FLOW

Electroosmosis is an electrokinetic phenomenon. When a current is applied to the microtube or microchannel a bulk flow of liquid occurs. In other words a liquid flows along a surface when an electric field is applied tangentially to the surface of the tube or channel. This flow is known as electroosmosis flow (EOF) and is a result of the surface charge on the inside surface of the channel or tube. The phenomenon can be explained as follows.

Most surfaces possess negative charges, which result from the ionization of the surface. These are called surface charges. A tube containing liquid now has two surfaces; the solid surface and the liquid surface. The immobile electric charges developed at the surface of a solid phase will in contact with the immobile charges developed at the surface of the liquid. Now an electric charge separation will occur at the interface of two-phases leading to the formation of an electric double layer (EDL). The double layers influence each other thereby causing the production of two layers, i.e. fixed layer and diffusion layer as shown in Fig. 10.5. In the diffusion layer the counter-ions (+ve ions) prevail over the negative ions. This in turn causes net local charge density to be nonzero. The electric potential developed because of the formation of EDL (note that all these occurs due to EDL only) is called *zeta* potential. An imaginary shear plane is formed at the vicinity of the fixed layer that separates the thin layer of liquid bound to the solid surface. The shear plane shows elastic behavior from the rest of liquids, which show normal viscous behavior.

When an electric field is applied parallel to the charged surface it will exert a Coulombic force on the ions within the electric double layer. The electric field makes it possible to build a pressure difference between the two ends of the tube, thereby causing them to flow in the direction determined by the direction of the electric field. The migration of the mobile ions will now carry the adjacent and hence the bulk liquid by viscosity. This flow is known as electroosmotic flow (EOF).

The net outcome of electroosmotic phenomenon is simply a pumping mechanism. The fluid can be transported to a microvessel through the capillary tube or a channel. The pressure-building ability of the electroosmosis has made it a powerful method to accurately transport and manipulate liquid samples in microfluidic devices for many applications such as chemical analyses and biomedical diagnoses. Since no movable part is involved in this mechanism, the system can have durable stability. Further,

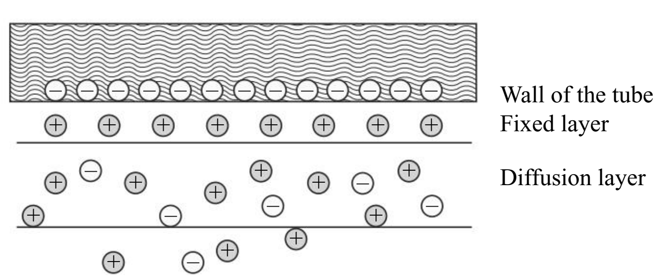


Fig. 10.5 *Formation of electric double layer (EDL) in electroosmosis flow (EOF)*

electroosmosis allows pumping of liquids over a wide range of materials. This is apparently essential for biochemical applications.

Mostly, DC currents are used, however, depending upon the requirement *AC* current can be used. In the latter case, as obvious, the flow is time and frequency dependent. Any change to the microchannel surface or the solution filling the channel will lead to changes in electroosmosis. Electroosmotic flow can be controlled by employing various methods. Table 10.1 shows methods to control EOF.

Table 10.1 *Methods to control EOF (Courtesy:* Agilent Technologies*)*

Variable	*Result*	*Comment*
Electric field	Proportional change in EOF	• Efficiency and resolution may decrease when lowered • Joule heating may result when increased
Buffer pH	EOF decreased at low pH, increased at high pH	• Best method to change EOF • May change charge or structure of solute
Ionic strength or buffer concentration	Decreases zeta potential and EOF when increased	• High ionic strength generates high currents and Joule heating • Low ionic strengths may result in sample adsorption • May distort peak shape if conductivity different from sample conductivity • Limits sample stacking if reduced
Temperature	Changes viscosity	• Often useful as temperature is controlled instrumentally
Organic modifier	Changes zeta potential and viscosity (usually decreases EOF)	• Complex changes, effects need to be determined experimentally • May alter selectivity
Surfactant	Adsorbs to capillary wall via hydrophobic and/or ionic interactions	• Anionic surfactants can increase EOF • Cationic can reverse or decrease EOF • Can significantly alter selectivity
Neutral hydrophilic polymer	Adsorbs to capillary wall via hydrophobic interactions	• Decreases EOF by shielding surface charge and increasing viscosity
Covalent coating	Chemical bonding to capillary wall	• Many modifications possible • Stability often problematic

10.12 OPTOELECTROWETTING (OEW)

Controlling surface tension by light is an interesting concept. Electrowetting and optoelectrowetting (OEW) methods are similar but in latter case a photoconductive layer plays an important role. Electrowetting is usually performed in an open surface channel whereas optoelectrowetting is performed in a closed rectangular channel. Figure 10.6(a) shows an open surface electrowetting method. For the purpose of understanding Fig. 10.6(b) also shows an open surface optoelectrowetting fluid flow setup. As can be seen that a photoconductive layer exist beneath the insulating layer. Figure 10.6(c) shows a practical closed rectangular channel with two layers of photoconductors. The liquid

droplet is sandwiched between the insulation layers integrated with two (top and bottom) photoconductive surfaces. Amorphous silicon and SiO_2 can be used as photoconductive material and insulation layer, respectively.

Much like electrowetting, the change in contact angle is determined by the voltage drop across the insulation layer. The electrical impedance of the photoconductor, insulator and droplet are found to be in series. The electrical equivalent circuit of one unit cell across a vertical cross section is shown in Fig. 10.6(d). Note that the channel is composed of many unit cells up to 25,000 depending upon the length of channel and the equivalent circuit can be a network as shown in Fig. 10.6(e). Consider only one unit cell. Within a unit cell various circuit elements are defined below.

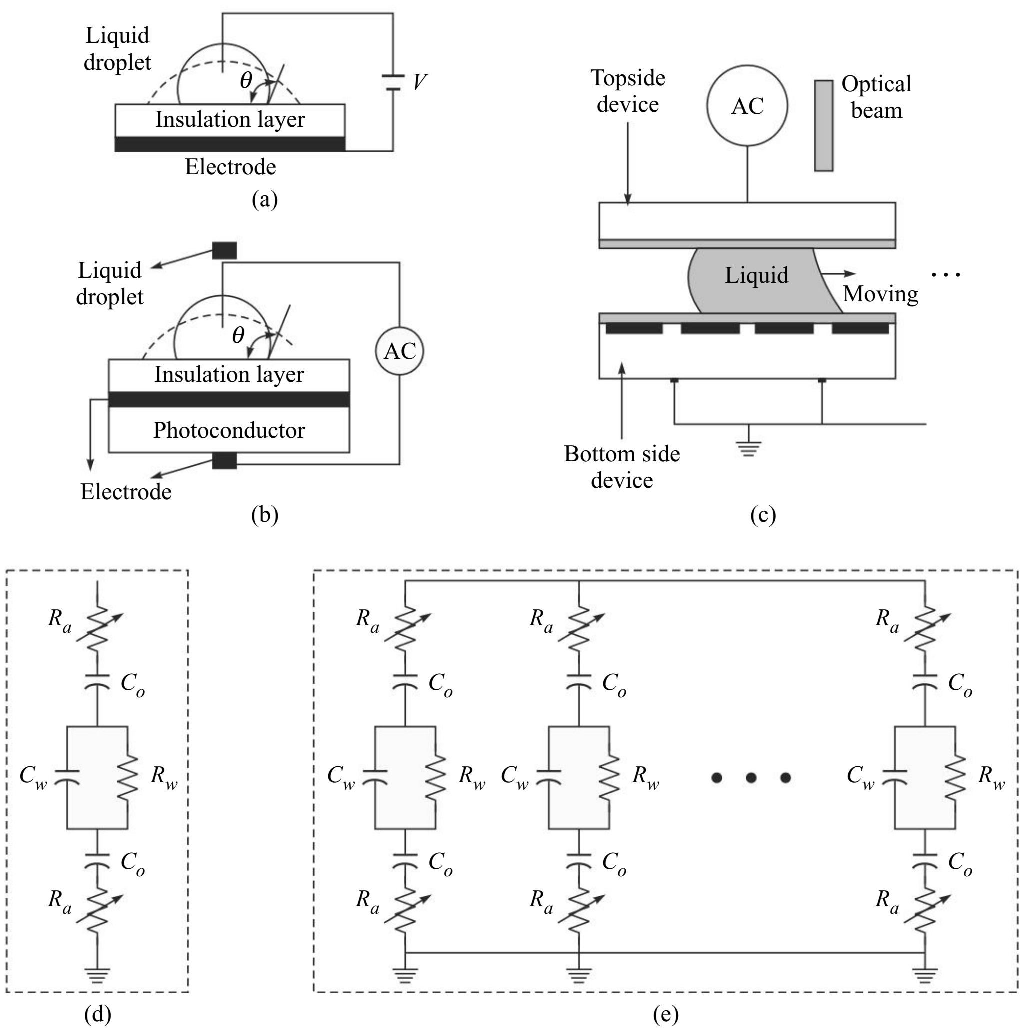


Fig. 10.6 *(a) Open surface dielectrowetting method, (b) Schematic illustration of open surface liquid droplets manipulation in opto-electrowetting (OEW) device, (c) Practical rectangular channel (d) Equivalent circuit of a single cell, (e) Equivalent circuit of the entire length of the channel*

R_a : Resistance of the amorphous silicon photoconductor under light illumination
C_o : Capacitance of the SiO_2 insulator
C_w : Capacitance of water layer between the two OEW surfaces
R_w : Resistance of water layer between the two OEW surfaces.

In a cell the voltage drop across the insulation layer is extremely small compared to photoconductor. That is most of the voltage drops across the photoconductor. When light falls vertically, the impedance of the photoconductor changes significantly and now most of the voltage drops across the insulator making a change in contact angle (Refer to Section 10.4.4). Eventually the contact angle is reduced by the illumination. In other words, in optoelectrowetting method the light changes the wettability factor of the photosensitive surfaces. The change in wettability factor makes actuation possible, i.e. fluid flow. By integrating a photoconductor in series with the electrowetting electrode, the voltage drop, V_i (Refer Eq. 10.13) across the insulator can be controlled optically. In practice, the photoconductor is actually integrated with the electrowetting electrode.

In typical applications, droplets from input reservoirs are injected and mixed by this method. It finally appears that projecting an optical beam on the photoconductor modifies the surface tension at the interface of the liquid and solid thereby allowing the liquid droplets to follow the optical beam path. The surface tension change is a reversible process. That is the surface changes from hydrophobic to hydrophilic when illuminated by light, and it returns to hydrophobic when light is removed. The experimental result shows that a fluid transportation speed up to 80 mm/sec can be achieved by OEW method.

The photoconductor is transparent and conductive. Amorphous silicon is mostly used as the photoconductive material. Amorphous silicon can be deposited by plasma-enhanced chemical vapor deposition (PECVD). Indium-tin-oxide (ITO) glass coated with 20 nm of Teflon behaves as the protecting seal. A two-dimensional array of electrowetting electrodes on a photoconductive material is integrated. Up to 25,000 numbers of electrodes can be integrated within the OEW structure. SiO_2 is mostly used as the insulating layer. The thickness of the SiO_2 is about 1 micrometer. The qualitative and quantitative parametric values of a typical OEW device is are given below.

- Size of the droplet should be bigger than the electrode
- Chip area is about 1 cm × 1 cm
- Area of the electrode is approximately 50 μm × 100 μm
- Number of electrodes = 20000–25000
- Contact angle = 100° (without illumination)–70° (with illumination)
- Operational frequency (f) = 100–600 Hz
- Applied AC voltage (V_{rms}) = 80–110 V
- Height of droplet is about 1–2 micrometer
- Incident light energy is about 4 mW at 530 nm wavelength
- Droplet speed varies from 10–80 mm per second

Injection, transportation, merging, and separation of nanovolume droplets actuated by optoelectrowetting-based scanning optical beams have precisely been experimented by many researchers. The OEW device requires only one common bias voltage for all electrodes, whereas in case of electrowetting technique this is not the situation. Here the electrodes are to be addressed optically by employing scanning principle. The integration of OEW components (electrodes, photoconductors and

insulators.) is a challenging problem, although the integration of large number electrodes is feasible. When manipulation of nanovolume or smaller liquid droplets is necessary the dimensions of the electrode areas can be scaled down. Simultaneous manipulation of multiple water droplets is also possible. The OEW method is better than the electrowetting method in the sense that the former has a good response time. Quantitatively, the OEW surface has a response time in the order of milliseconds. Other advantages are that the OEW devices are flexible and reconfigurable. The methodology can be used to detect the species by fluorescent detection and after the detection the drop can be sorted in a reservoir using optical beams for further mixing or processing.

10.13 TUNING USING MICROFLUIDICS

In optical communication, tuning of physical properties of the optical fiber is required in many situations. Researchers at Lucent Technologies and Bell Laboratories have developed a microfluidic device that allows the tuning of properties of the optical fiber. The developed device can tune the properties replacing the traditional methods of tuning such as movement of parts and thermal control. The method involves manipulating of liquid inside an optical fiber. The cladding of the fiber has a series of holes encircling the core. The holes can be filled up with liquid. By the use of built-in heaters, liquid of a high refractive index can be moved to a desirable sensitive section of the fiber, which in turn can change the properties. Figure 10.7 (top) shows the cross section of a tunable microchannel fiberoptic device. The device has two sections, namely the microchannel and an upper glass lid and pump (bottom left).

10.14 TYPICAL MICROFLUIDIC CHANNEL

Figure 10.8 illustrates a typical microfluidic channel, which consists of a surface micro-machined labyrinth, having one central inlet and two outlets. This device imported from Ansys CAD software, is a model, which is approximately 100×120 μm dimension, with a channel depth of 10 μm. The top of this device is not shown because of clarity, but is made by sealing another layer onto the top surface. Fusion bonding (see Chapter 2) technique can be employed to join the channel and the top layer. The labyrinth can function as a pressure drop or pulse attenuator for blood flow and can be used in clinical diagnosis. Blood can flow into the central inlet and out through the two flanking outlets at a reduced pressure. Figure 10.8(c) shows the streamline coded in pressure (Central input channel = high pressure; Two output channel = low pressure), plus particles that follow the streamlines. Particle length is proportional to fluid velocity. An approximate Guassian velocity distribution is shown across the width of the channel, highest velocity in the center of the channel. In essence, the device can be utilized for the purpose of liquid analysis such as:

- Determination of pressure drop across the device
- Determination of velocity profile
- Computation of pressure applied to walls
- Transfer of heat from fluid to structure and vice-versa.

Obviously, the above liquid (blood) parameters within the channel can be analysed.

10.15 MICROFLUID DISPENSER

Researchers have reported a highly parallel nanoliter dispenser, which is fabricated by high-speed micromilling in different plastic materials. The device consists of up to 1,536 microstructured

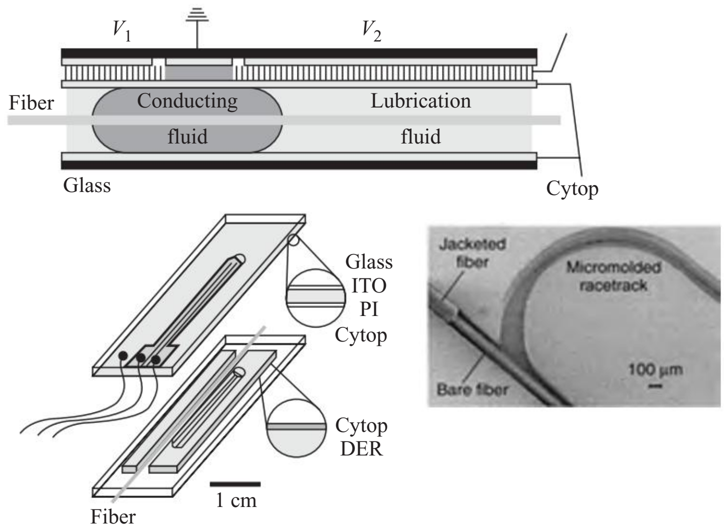


Fig. 10.7 *Simple microfluidic system tunes fiber properties*
(Source: LaserFocusWorld, 2003; Courtesy: LUCENT TECHNOLOGIES*)*

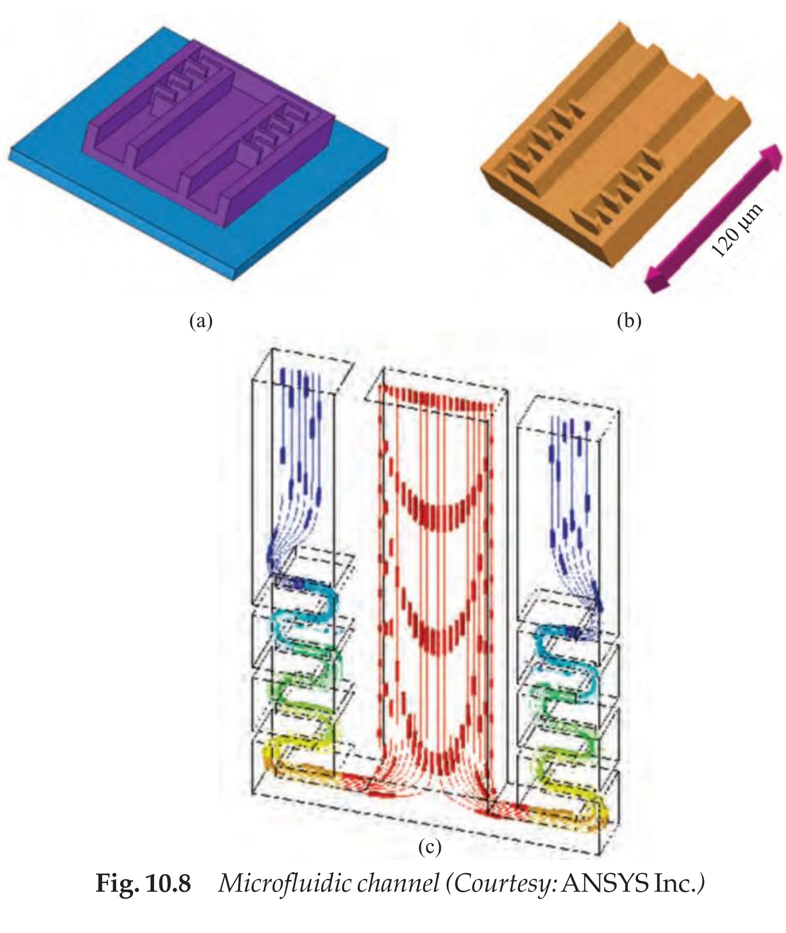


Fig. 10.8 *Microfluidic channel (Courtesy:* ANSYS Inc.*)*

dispensing units arrayed on a macroscopic area in the format of a standard micro well plate (approximately 81 × 123 mm). Each dispensing unit consists of 3 basic elements: a reservoir, a connection channel and a tapered nozzle. The challenge has been to fabricate these elements in a real 3D-geometry on a 1 mm thick substrate by high-speed micromilling process.

10.16 MICRONEEDLE

Microfluidic systems promise to revolutionize health care by providing equipments for precise delivery and control of biological fluids. To achieve this in many situations, microneedles are used. Microneedles are attractive from a design perspective as they are also compatible with MEMS fabrication process. Because of their small size they can be fabricated to provide a range of geometries and flow characteristics. Microneedles (Fig. 10.9) are considered as an important BioMEMS (see Chapter 11) devices and especially very useful for the following applications.

- Collection of samples for biological analysis
- Delivery of cell or cellular extract based vaccines
- Providing interconnection between the microscopic and macroscopic devices
- Extra and Intracellular neuronal recording

Microneedles are continuous low-flow-rate based drug delivery devices. An immediate example of continuous low-flow-rate delivery application is the injection of insulin to a diabetic patient. The research team at Berkeley Sensor and Actuator Center at University of California have designed three types of microneedles whose dimensions, specification and performances are given in Table 10.2(a) and (b). For a laminar flow, the average flow rate Q can be expressed as,

$$Q = \frac{4ba^3}{3\mu}\left(-\frac{dP}{dx}\right)\left[1 - \frac{192a}{\pi^5 b}\sum_{i=1,3,5\ldots}^{\infty}\frac{\tan h(i\pi b/2a)}{i^5}\right] \tag{10.18}$$

which is obtained by integrating x-directed velocity profile of a rectangular duct with y and z as cross section. Here $2a$ is the length of the walls, $2b$ the length of other wall, dP/dx is the pressure gradient which can be estimated to be,

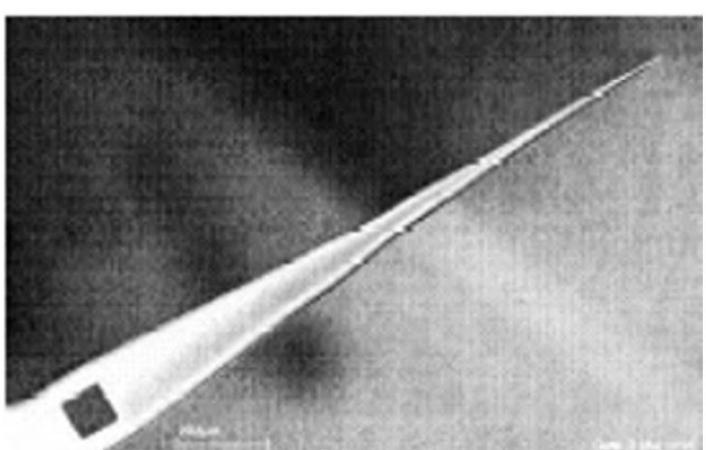


Fig. 10.9 *Microneedle (Source:* Judy, University of California*)*

$$\frac{dP}{dx} = -\frac{4\tau_s}{D}, \text{ where } \tau_s = \frac{0.332\mu U}{x}\sqrt{\Re_x} \tag{10.19}$$

τ_s is shear stress of the plate, x is the distance along the plate, $\Re_x$ is the Reynolds number based on the distance x. The above expression is based on the assumption that the rectangular fluid duct behaves as a collection of four places. The average velocity can be expressed as

$$U = \frac{Q}{4\,ab} \tag{10.20}$$

Table 10.2 (a) *Neddle parameter*

Dimension in micron	*Bent needle*	*Straight needle*	*Filter needle (before/after the experiment)*
Length	5500	5000	5000/1000
Width	160	80	80/1350
Height	80	80	60/60
Hydraulic diameter	170	80	60/68.6

Table 10.2 (b) *Experimental vs. Analytical flow rate at a constant pressure head 138 kPa*

Needle type	*Number of tests*	*Avg. flow rate (ml/s)*	*Analytical flow rate*	*Error (%)*	*Reynolds number*
Bent	4	0.082	0.088	7.3	738
Reinforced	9	0.040	0.040	0.0	503
Filter	2	0.070	0.083	17.9	688
Double channel	1	0.032	0.034	6.2	260

10.17 MOLECULAR GATE

The molecular gate is understood as a nano level construct that interconnects microfluidic channels with the nanopore membranes. Analogous to the principle of operation of a semiconductor transistor, from digital point of view, the molecular gate can also digitally deliver and control the amount of materials in the order of sub-attomoles by applying electric potentials. It is understood that a semiconductor device such as diode delivers and controls electrons. Similarly the molecular gate can deliver and control the material but with added functions. Additional functions are due to the reason that the molecules undergo chemical reactions, changing both the composition and behavior of the fluid.

A very approachable potential application of these molecular gates is to control and deliver mass-limited proteins and even pharmaceuticals in precise amounts such as attoliter volumes and attomolar concentrations. Moreover, filtering action (Fig. 10.10) is also achieved. The undesired biomolecule are filtered out based on their size, mass and chemical affinity and desirable biomolecule are allowed to pass through. The molecular gate is known as molecular delivery system and mostly used in medical applications. For instance, the concentration of a variety of target metabolites in serum can be known in order to dynamically respond to the desired compound to be delivered, thereby improving the health of the person.

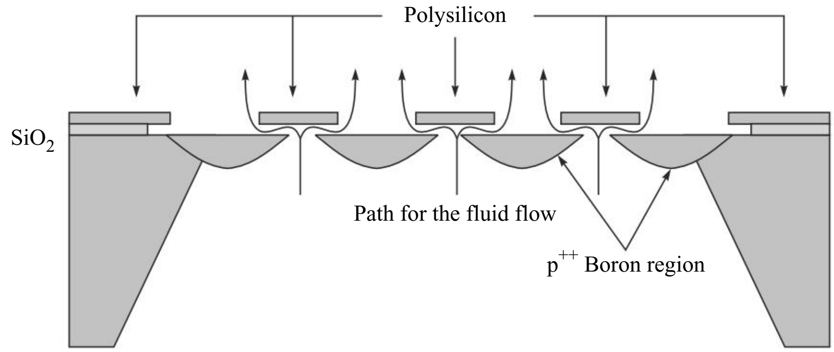


Fig. 10.10 *Filter (Source: Judy, University of California)*

With regard to fabrication, silicon is used as the substrate. p^{++} boron doping is performed by using thermal diffusion method. A thin sacrificial layer is deposited on the diffused boron surface. Polysilicon is then deposited on the top of the sacrificial layer. By using photolithography and etching process the fluid flow paths are created.

10.18 MICROPUMPS: THE CONTINUOUS FLOW SYSTEM

A mechanical machine that moves fluid or gas continuously by suction pressure is known as a pump. Water pumps are common in our everyday lives. However, micropumps are MEMS devices, which are primarily used for microfluidic applications. The cost-effective transport of small quantities of biochemical fluidic samples, in the range of microliters per minute, has been an important challenge for micropumps. Typically, these devices operate with flow rates in the range of nanoliter to microliter per minute. Various types of MEMS-based micropumps exist. Diaphragm based micropumps shown in Fig. 10.11 are common. The diaphragm is driven back and forth to achieve pumping action. Diaphragm-based micropumps are capable of providing good performance.

10.18.1 Construction

The important part of the micropumps is the diaphragm. The back and forth operation of the diaphragm is achieved by applying AC signal. The material should have high electromechanical properties such as (i) high electrostrictive strain, (ii) high energy density, and (iii) high displacement voltage ratio. Polymer is chosen as the suitable material. Mostly, vinylidene fluoride-trifluoroethylene polymers are used, as they possess these properties. These are called high-energy electron irradiated poly. Besides diaphragm, the other parts and sections of the micropumps are listed below.

- Counter electrode
- Isolation layer
- Actuation chamber
- Pump chamber
- Inlet valve
- Outlet valve
- Inlet microtube
- Outlet microtube

As mentioned the pump is electrically driven. The AC supply with appropriate frequency is applied across the two electrodes. The diaphragm itself constitutes one electrode. The other terminal of the AC supply is applied to the counter electrode. Isolation layer is a design criterion and it has high impedance.

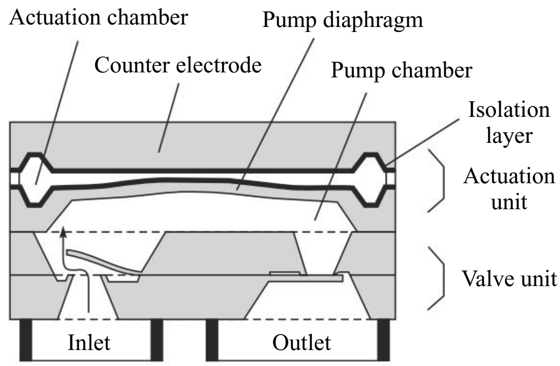


Fig. 10.11 *3D cross-sectional view of a diaphragm-based micropump assembly (Source: Judy, University of California)*

When in action, the layers (top and bottom) can protect both the electrodes in case of probable contacts that may arise because of accidental high input supply voltage. Actuation chamber, groove-like opening, encircling the circular diaphragm facilitates flexibility as far as back and forth movement of the diaphragm is concerned. Further, the chamber creates more space and helps in accommodating the air when the diaphragm moves upward and downward. The diaphragm, actuator chamber and the gap between the electrodes constitute the *actuation unit*. Pump chamber is a section through which the fluid is moved from the inlet to the outlet. There are two valves, called inlet valve and outlet valve, interfacing the inlet and outlet microtubes, respectively. The valves are unidirectional and are called check valves. Inlet valve allows fluid to enter into the pump chamber whereas the outlet valve allows the fluid to move away from the chamber. Flow is accomplished by using these one-way check valves. Each valve has a valve cap that is tethered over an orifice to a silicon substrate. The valve, pump chamber and the orifice constitute *valve unit*. The materials used for fabricating the valves are mostly parylene. Microtubes must be there to facilitate transportation of fluid from the source to the destination.

10.18.2 Principle of Operation

The operation of the micropump is shown in Fig. 10.11. The figure shows 3D cross-sectional view of a pump assembly. By applying AC voltage the diaphragm can be deflected up and down (back-and-forth). During positive cycle of the applied signal, the diaphragm makes upward movement. Because of this a suction pressure is developed within the pump chamber. The pressure makes it possible to open the cap of the inlet valve and to allow fluids to enter into the pump chamber. During this half-cycle the cap of the outlet valve is closed, as the cap exists at the other end of the orifice. During negative excursion of the applied AC signal the diaphragm makes downward movement causing the cap of the outlet valve to open. The downward movement makes it possible to allow the fluid to pump away from the chamber.

10.18.3 Modeling

The pump can be modeled by considering the diaphragm as a circular plate. It is assumed that the circular plate with clamped edges is subjected to uniform mechanical pressure in the lateral direction. Let the radius of the plate be R_0 and the horizontal and vertical axes are denoted as r and z, respectively as shown in Fig. 10.12.

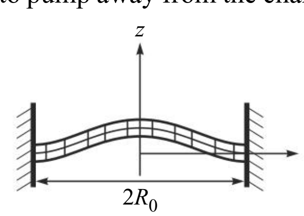


Fig. 10.12 *Deflection of diaphragm under an electric field*

When an electric field is applied to the plate the diaphragm will be displaced as shown. Now, the displacement along the r-direction is expressed as follows.

$$z = c(R_0^2 - r^2)^2 \tag{10.21}$$

where, c is a constant related to transverse strain. From the above expression, dz can be obtained as,

$$dz = 4c(R_0^2 - r^2)rdr \tag{10.22}$$

The change in volume V, within the pump chamber is given by,

$$V = \int_0^{z_0} \pi r^2 \, dz \tag{10.23}$$

where z_0 is the displacement at the center. Putting dz from Eq.10.22 in Eq.10.23 and changing the limits from z to R, the above equation can be written as,

$$V = 4\pi c \int_0^{R_0} (R_0^2 - r^2)^3 \, dr = \frac{1}{3} \pi c R_0^6 \tag{10.24}$$

The volume rate or pumping speed (expressed in terms of microliter per minute) at the atmospheric pressure can be expressed as

$$V_s = 60 \times f_d V = 20\pi c R_0^6 \tag{10.25}$$

where f_d is the frequency of diaphragm. Typical values of pumping rate are 50–70 microliter per minute at the applied voltage of 2–3 V_{pp}. The pump pressure can be about 600 *Pa* at this voltage. The operational frequency ranges from 10–30 Hz. The overall dimensions of MEMS micropump can be approximately 5000 × 5000 × 1000 μm with respect to the length, width and height, respectively.

Other driving methods such as electromagnetic, electrothermal etc. can be employed to vibrate the diaphragm. Table 10.3 shows a performance comparison of the micropumps actuated by other driving method.

Table 10.3 *Performance comparison of the micropumps (Source 0-7803-5998-4/01/@2001 IEEE)*

Actuation Type	*Maximum flow rate (• l/min)*	*Operation voltage (V)*	*Power consumption (mW)*	*Maximum pump Pressure (kPa)*
Piezoelectric	1300	160	-	90
	40	100	-	15
Electrostatic	850	200	1	31
Thermopneumatic	34	6	2000	4
Electromagnetic	20	3	900	-

10.19 MICROFLUIDIC DESIGN CONSIDERATIONS

Three important design parameters are considered while designing the microfluidic systems. They are

- Preparation of suitable surfaces
- Well-defined pathways for the liquid to flow
- Appropriate selection of materials

In almost all microfluidic applications such as chemical analysis and synthesis, for instance, the liquids are mostly aqueous solutions. For the transportation of these solutions the desired surfaces and

pathways are hydrophobic and hydrophilic. The hydrophobic surface consists of thin layer applied to the surface via dip coating or spin coating. In many cases self-assembled monolayers can also be used. Most of the MFD are fabricated in glass or silica or silicon based substrates. The hydrophilic areas are called wettability patterns and are usually produced by three methods such as photolithography, local plasma treatment, or by microcontact printing.

Photolithographic processing techniques are commonly used to develop channels in the surface of a planar substrate that are then covered with a thin oxide layer of similar material. The advantage of using these materials is that their *electrophoretic* and *chromatographic*[3] properties and surface *derivatization* chemistries are extensively characterized and best understood. Other materials including quartz, those etch more slowly are also used to develop channels. The channels etched into glass or quartz is sealed by high-temperature annealing process. Microfluid devices prepared entirely in silicon are sealed by wafer bonding.

Contrast to silicon-based substrates, polymer substrates are also promising alternatives. The interest in using plastics is primarily driven by the fact that these materials are less expensive and easier to manipulate than silica-based substrates. Further, plastics are well suited to casting, molding, laser ablation, and machining operations.

10.20 SUMMARY

The study of flow of microvolume liquid through microdevice is called microfluidics. A microfluidic device is a structure fabricated in silicon, glass, or polymer on a chip. It can also allow the flow of gases that can be manipulated and controlled through valves and switches. The microfluidic device or system is essentially instrumental in detection, analysis, sample preparation and delivery of a wide range of chemicals, drugs, biological solutions, etc. in industry as well as biomedical applications. Microfluidics is playing a major role in diagnosing and treating a number of diseases and conditions. Microfluidic channels in MEMS devices are playing a key role for the next generation biomedical applications. The key to the microfluidic system design is the understanding of the principle of operation of transportation of fluid through the microchannel.

Because of the different physical properties of fluids at the microscale—for instance in channels of sub-millimeter diameter, fluid flow is not very turbulent but always nearly laminar, microfluidics include exclusive knowledge on the flow phenomenon as well as on the design and fabrication of miniaturized device. Since you have already gathered knowledge on the design aspects of microsystems, we preferred to present the principles of various flow phenomena in this chapter. In particular, this chapter was dedicated to describe the digital flow methods, which can be achieved through Dielectrophoresis (DEP), Electrowetting, Electrothermal, Thermocapillary and Electroosmosis principle. At the beginning the chapter describes the properties of fluid.

The continuous flow of fluid is achieved by the use of micropump. Last part of the chapter describes the design, construction and modeling of a diaphragm-based MEMS micropump. Further, a molecular gate, which can be used for filtering application, was described.

[3] A process of separating the small quantities of substance of a mixture on the basis of differences in their affinity for a stationary and a mobile phase.

Points to Remember

- The study of motion or transportation of fluids and their mixtures at a microscale level is known as microfluidics.
- Microdevices which are used to transport and store fluid are called microfluidic systems (MFS) or microfluidic devices (MFD).
- Microfluidics leads to the establishment of a foundation knowledge on modeling, developing and designing typical microfluidic devices such as the microcapillary tubes, channels and systems which are eventually used in many fields like life sciences, chemistry, medicine, computing (inkjet printing), etc.
- Microfluidics implicitly concerns the design and development of microscale devices for medical, chemical and biotechnological research. These devices are rather called tools, which have been emerged as biochip or lab-on-a-chip.
- The surface tension provides the required wall tension in order to form a bubble.
- A droplet tends to form a globular shape by the cohesive forces of the surface layer.
- A fluid is defined as a substance that cannot support shear forces in steady state.
- Fluid flow is characterized by laminar flow and turbulent flow.
- The laminar flow is characterized by low Reynolds number signifying the flow is dominated by viscous forces, whereas the Reynolds number is high in case of turbulent flow and the typical value is more than 2000.
- Molecules on the surface, however, experience an imbalance of forces. Since the molecules on the surface of a liquid experience an imbalance of forces because of missing bonds of some sort, the net effect of this situation is the storage of free energy at the surface of the liquid. Surface tension tries to minimize the surface area, forming spherical droplets.
- The effect of microfluidic surface tension can be utilized in many applications. The effect can be employed to act as the medium of power transmission.
- The incompressible, quasi-steady, laminar and Newtonian *2D* horizontal flow of liquid is expressed as what is known as Navier-Stokes equation.
- The solution to momentum and continuity equation through computer is referred to as Computational Fluid Dynamics (CFD). Apparently, CFD is a computational technology that enables to study the dynamics of things that flow.
- The microfluidic systems are of two types based on the way the microvolume fluid is transported or its position is manipulated: Continuous flow systems and Droplet-based flow system.
- When the systems uses surface tension gradient to move, combine, and mix liquid droplets, the transportation method is called *digital* fluidic microsystem.
- Some of the important digital mechanisms are, Dielectrophoresis, Electrowetting, Electrothermal, Thermocapillary, Electroosmosis and Optoelectrowetting.
- When a potential difference between the two electrodes is applied a force is exerted on the liquid, trying to impel in a specific direction. This process is called Dielectrophoresis.
- The principle behind the electrowetting method is that the contact angle of a partially wetting conductive liquid on an insulating substrate can be reduced by applying a voltage between the liquid and a counter electrode sitting underneath the insulating layer.
- Electrothermal flow arises from the temperature gradient in the medium generated by joule heating of the fluid. Because of the gradient the parameters such as conductivity, permittivity, viscosity, and density of the liquid are changed, establishing gradients in these parameters. The gradients then make it possible to generate forces, which eventually act on the fluid.
- There is a similarity between the electrothermal and thermocapillary flow however, in the latter case the effect is on a capillary tube.

- Whenever fluid dimensions are less than 10 microns, thermocapillary-driven forces makes it possible to separate the droplets discreetly from the main fluid body contained in the microcapillary tube.
- A liquid flows along a surface when an electric field is applied tangentially to the surface of the tube or channel.
- This flow is known as electroosmosis flow (EOF) and is a result of the surface charge on the inside of the wall.
- Electric charge separation will occur at the interface of two-phases leading to the formation of an electric double layer (EDL). The double layers influence each other thereby causing to produce two layers such as fixed layer and diffusion layer.
- The pressure-building ability of the electroosmosis has made it a powerful method to accurately transport and manipulate the liquid samples in microfluidic devices for many applications such as chemical analyses and biomedical diagnoses.
- Controlling surface tension by light is called Optoelectrowetting (OEW).
- Electrowetting is usually performed in an open surface channel whereas optoelectrowetting is performed in a closed rectangular channel.
- In optical communication, tuning of physical properties of the optical fiber is required in many applications. A microfluidic device can allow tuning of properties of the optical fiber.
- Microneedles can be used for continuous low flow rate based drug delivery applications.
- The molecular gate is understood as a nano level construct that interconnects microfluidic channels with the nanopore membranes. Analogous to the principle of operation of a semiconductor transistor, from digital point of view, the molecular gate can also digitally deliver and control the amount of materials in the order of sub-attomoles by applying electric potentials.
- A mechanical machine that moves fluid or gas continuously, unless until stopped, by suction pressure is known as a pump.
- Various types of MEMS-based micropumps exist. Diaphragm based micropumps are common.
- The important part of the micropumps is the diaphragm. The back and forth operation of the diaphragm is achieved by applying AC signal. The back and forth movement is twice that of the frequency of the supplied AC signal.
- Actuation chamber, groove-like opening encircling the circular diaphragm facilitates flexibility as far as back and forth movement of the diaphragm is concerned.

Exercises

1. What do you understand by macrofluidics and microfluidics?
2. What are the usefulness and features of microfluidic devices?
3. Why is fluid analysis made? In which fields is fluid analysis illustrious? What are the main advantages of using microscale fluids for analysis?
4. What do you mean by lab-on-a-chip (loac) device? Is this chip an IC chip?
5. The behavior of fluid is significantly changed as geometric scale decreases. In this respect some considerations are important to both the users and system designers. What are those considerations?
6. Define the following laws and terms.
   (a) Laplace Law
   (b) Static fluid pressure
   (c) Energy density
   (d) Bubbles and droplets

7. Write notes on the following.
   (a) Density of fluid
   (b) Viscosity
   (c) Surface tension
8. Write down the analytical expressions of liquid flow in a channel.
9. Express the meaning of CFD. Why is CFD important?
10. Differentiate between continuous and digital microfluidc flow. What are the various mechanisms employed for achieving the digital flow?
11. With the suitable diagrams, explain the principles of the following fluid flow phenomena, respectively.
    (a) Dielectrophoresis (DEP)
    (b) Electrowetting based fluid flow
    (c) Electrothermal flow
    (d) Thermocapillary based fluid flow
    (e) Electroosmosis flow
    (f) Optoelectrowetting
12. Write down the relationship between the contact angle and the applied voltage in case of electrowetting based fluid flow. Define the terms.
13. What do you mean by the term tuning and how can the properties of the optical fiber be tuned by using microfluidic systems?
14. With neat schematic diagram explain the role of a typical molecular gate.
15. How can continuous fluid flow be achieved by using a micropump? What are the important parts of a typical micropump? Draw the schematic diagram of a diaphragm based micropump and explain its principle of operation. Further, model this micropump in order to express the volume rate or pumping speed at the atmospheric pressure.
16. Prepare a performance table in order to show the maximum flow rate and operational voltage ranges of the piezoelectric, electrostatic, thermopneumatic and electromagnetic actuation type micropumps, respectively.
17. What are the three important parameters, which are considered while designing the microfluidic systems?

Chapter 11

Chemical and Biomedical Microsystems

Objectives

The objective of this chapter is to study the following.

- Biomedical and chemical microdevices
- BioMEMS and Chemodevices
- Chemo resistor, chemocapacitor and chemotransistor
- DNA (Deoxyribonucleine-Acid) sensor
- Surface Acoustic Wave (SAW) sensors

11.1 INTRODUCTION

The scope of MEMS technology has recently been extended to health sciences and chemical industry. MEMS devices used for health science and chemical industry are called BioMEMS and chemical microsystems. Both BioMEMS and chemical microsystems are commonly called *biochemical microsystems* (BCMS). In other words, BCMS is defined as the methods and principles that allow the development of biochemical sensitive microdevices. Areas of study and development include diagnostic tools, genomics, drug discovery and drug delivery systems. Some important designs are chemosensors, pollution and odor detectors (e-nose) and DNA (Deoxyribonucleic Acid) analyzers (An introduction to DNA is given in Appendix C). BCMS can be classified under two topics; BCMS actuators and equipments (BCMSAE) and BCMS sensory devices (BCMSSD). BCMSAE deals with the study of surgical equipments, instruments and micro artificial organs. On the other hand, BCMSSD includes fundamental study of sensory devices based on various transduction principles. This chapter deals with BCMSSD. The development of BCMSAE is still in the rudimentary stage of research. However, basic operational principles of some of these systems have already been presented in the previous chapters (In Chapter 8—Magnetic actuator for microsurgery applications and in Chapter 10—Microneedle).

11.2 SENSING MECHANISM

The sensing mechanism in BCMSSD is popularly known as biological and chemical detection method. The sensor is capable of converting biological or chemical quantity into an electrical signal (Voltage or current). The basic transduction structure it includes is a chemically sensitive layer followed by a transducer element as shown in Fig. 11.l. In many cases, the sensitive layer itself behaves as the transducer element. The sensitive layer is simply a membrane. Upon interaction with a chemical species called analytes the physical, electrical or optical properties of the membrane changes. The properties can be;

- Volume or Mass
- Resistance
- Optical properties

The interaction can be in the form of absorption, chemical reaction and charge transfer. The change is reversible. That is the membrane can restore the original property when interaction is inhibited. The role of transducer is to detect the change that occurs in the sensitive membrane. The transducer output can be an electrical signal such as current or voltage or even frequency. The output can then be amplified, calibrated and displayed or read out. Sometimes the output is subjected to further data processing in order to eliminate and compensate noise and error, respectively.

From the designer's perspective, the following points draw attention while designing the sensory devices. The rest of the chapter deals with these factors in more detail.

- Stimulus of measurement
- Physical phenomenon that it is sensitive to
- Conversion mechanism
- Material for fabrication
- Field of application

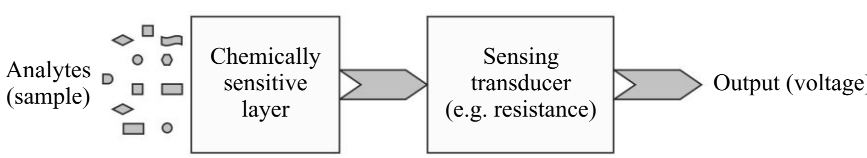


Fig. 11.1 *Principle of operation of biochemical sensors*

11.3 PRIMARY SENSING PRINCIPLE

The BCMS sensors are batch-processed miniaturized systems, optimally linked together in order to achieve specific mechanical, electrical or chemical functions. There are various types of transducers based on different physical principles. Table 11.1 lists out some of the important sensory devices and corresponding affected parameters and principles. Among them the following three principles are commonly employed in sensor design.

- Electrochemical sensors
- Chemomechanical sensors
- Thermochemical sensors

A classification in this respect is given in Fig. 11.2. Electrochemical sensors change the electrical property such as resistance and potential. Chemomechanical sensors are called mass-sensitive sensors

due to the reason that the physical property such as mass of the changes upon absorption. In thermal sensors, the temperature changes through chemical interaction. Electrochemical and chemomechanical sensors are common.

Table 11.1 *Sensory devices and their corresponding affected parameters and principles*

Principle	*Change in parameter*	*Transducer*
Calorimetric	Temperature	CalSpec
Capacitive	Capacitive or charge	Chemocapacitor (e.g. Polymer pH sensor)
Conductive	Resistance	Chemoresistor (e.g. SnO_2 gas sensor)
Fluorescent	Intensity	Fiber waveguide
Gravimetric	Mass	Piezoelectric
Optical	Path length	IR detector (e.g. Methone detection)
Potentiometric	Potential difference	Ion selective FET
Resonant	Frequency	SAW sensors

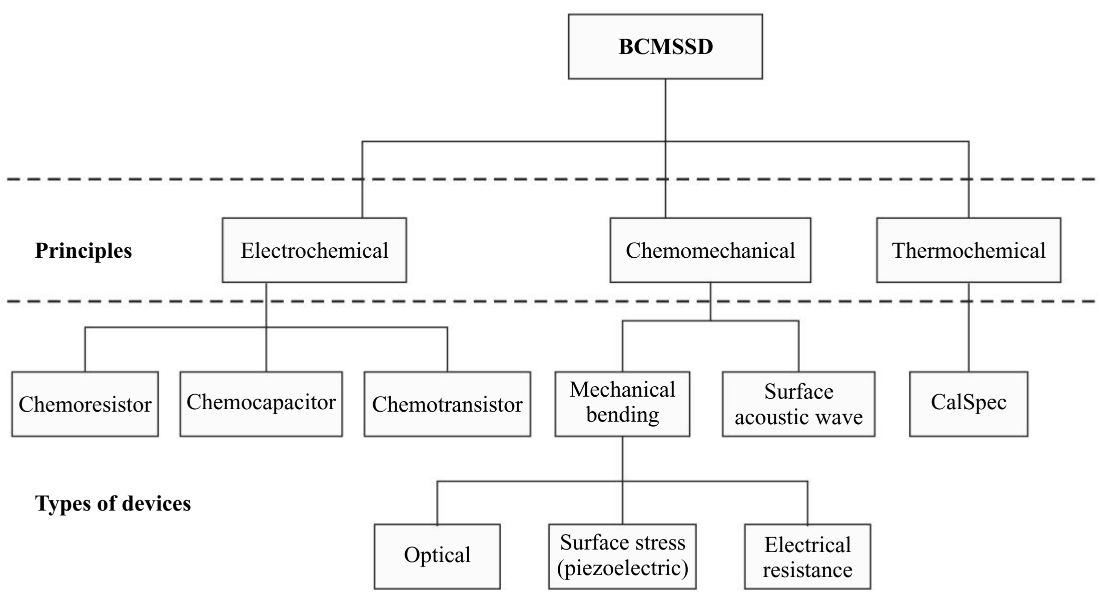


Fig. 11.2 *Principle of operation of BioMEMS and Chemodevices*

Regardless of their transduction principle, all sensors are always characterized by sensitivity, stability, applicability, portability and cost. Further, the sensors are to produce either discrete or continuous measurand, which are proportional to a single analyte or a group of analytes. Sensors for use with biological applications require low operating temperatures and biocompatibility.

11.4 MEMBRANE-TRANSDUCER MATERIALS

The chemically sensitive coated layers can be organic or inorganic materials. They are coated to respond to different chemicals. Organic layers are either conducting or nonconducting polymers. Some

of them are polysiloxanes, polyurethanes or polyaniline. These polymers can be used for monitoring hydrocarbons and halogenated compounds. Polymer layers can also detect toxic volatile organics. The sensitive materials and their operation conditions (e.g. temperature) impose certain requirements on the transducer design. The inorganic membrane materials are electron-conducting oxides such as tin dioxide (SnO_2) and zirconium dioxide (ZrO_2). SnO_2 is used for the detection of hydrogen, carbon monoxide, and nitrogen oxides. ZrO_2 is used to determine oxygen, nitrogen oxide and ammonia. Some important materials are given in Appendix B.

Shape and coating sensitive silicon chemical sensing devices are commonly used to detect combustible gases such as CO, CO_2, NO_x, N_2, and even H_2O. The conventional devices are being replaced by these micro shape and coating sensitive silicon chemical sensing platforms because the former devices dissipate power in the form of heat. Silicon wafer based device has a substrate frame upon which a dielectric membrane is suspended. A thin platinum film is built into the membrane. The film serves as the heater for the membrane. Due to the presence of combustible gases a coating will be formed thereby generating heat causing the membrane to reshape. The change of the geometry and the coating of the suspended mass are optimized to detect CO, CO_2, NO_x, N_2 and H_2O. The dielectric membrane on a silicon frame achieves thermal isolation signifying less power dissipation. The detection of these gases is carried out in many sectors such as industrial, automotive, laboratory and health diagnostic instruments.

11.5 CHEM-LAB-ON-A-CHIP (CLOC)

BioMEMS and chemical microsystems (BCMS) are also used in microfluidic applications. For specific application such as pollution, chemical and DNA analysis both BCMS and MFS (Microfluidic Systems) are sometimes integrated on one platform. A typical integrated platform may include tiny channels, which transport and control the flow of microvolume of gases and liquids for further detection, analysis and processing. These integrated devices are called portable laboratories in which analyses are performed. These small-sized portable laboratories are known as *chem-labs-on-a-chip* (CLOC). Normal photolithography and etching techniques are used to fabricate the CLOC devices. However, special attention is paid as far as selection of biocompatible materials are concerned.

11.6 CHEMORESISTORS

Chemoresistor is simply a transducer device in which the resistance of a chemically sensitive layer changes with respect to the amount of chemical absorbed. Usually the chemoresistors are used for the detection of gases. The variation of the resistance in the presence of a gas can be expressed as follows.

$$R = R_0\left(1 + kC^{\gamma} e^{\frac{t_c}{T}}\right) \tag{11.1}$$

where R_0 is the resistance when gas exposure is zero; k is called coefficient of sensitivity of the resistive material; C^{γ} is the gas concentration (expressed in ppm[1]); t_c is a temperature coefficient and T is the temperature in $^{\circ}$K. Depending on the material used for the transducer the coefficient of sensitivity can either be positive or negative, thereby producing an increase or decrease in the sensor resistance once the gas is exposed. Note that an increase in temperature results in decrease in the sensor resistance when K_S is positive. A modified version of the above equation is used to measure humidity.

[1] ppm (parts per million) is a ratio figure that represents the amount of one substance that is in one million parts of another substance.

$$R = R_0\left(1 + k_g C_g^{\gamma_g} e^{\frac{t_{cg}}{T}} + k_w C_w^{\gamma_w} e^{\frac{t_{cw}}{T}}\right) \tag{11.2}$$

where, the subscripts g and w correspond to gas and water vapor, respectively.

The schematic diagram of a chemoresistor is shown in the Fig. 11.3. The device has two terminals, which internally connect to two separate electrodes. In between the electrodes active material, the main ingredient of the chemoresistor exists. The circuit current and hence the voltage can be measured to detect the resistance and thereby the gas. The electrodes and active materials are deposited by usual photolithographic fabrication process. Silicon, SiO_2 or AlO can be used as the substrate. Metal oxide or organic materials such as organic crystals and conducting polymers are taken as the active materials. When metal oxide is used, the device is called inorganic gas sensor. Some of the metal oxides are ZnO, TiO_2, In_2O_3 and SnO_2. On the other hand, when organic crystal or conducting polymers are used, the sensors are called organic gas sensors. The organic material, especially polymers are very sensitive compared to inorganic sensors. The major advantage of the use of polymers as active material are the excellent linearity of the transduced signal as a result of the absorption of the biomolecules to be detected and their long-term stability. Polymers also characterize good response time, selectivity and reversibility. Response time of a material is defined as its ability to produce output with respect to time parameter. Selectivity is defended as the ability to produce output with respect to a fixed input. In other words, selectivity is the degree to which the interaction produces the desired result. A material with high

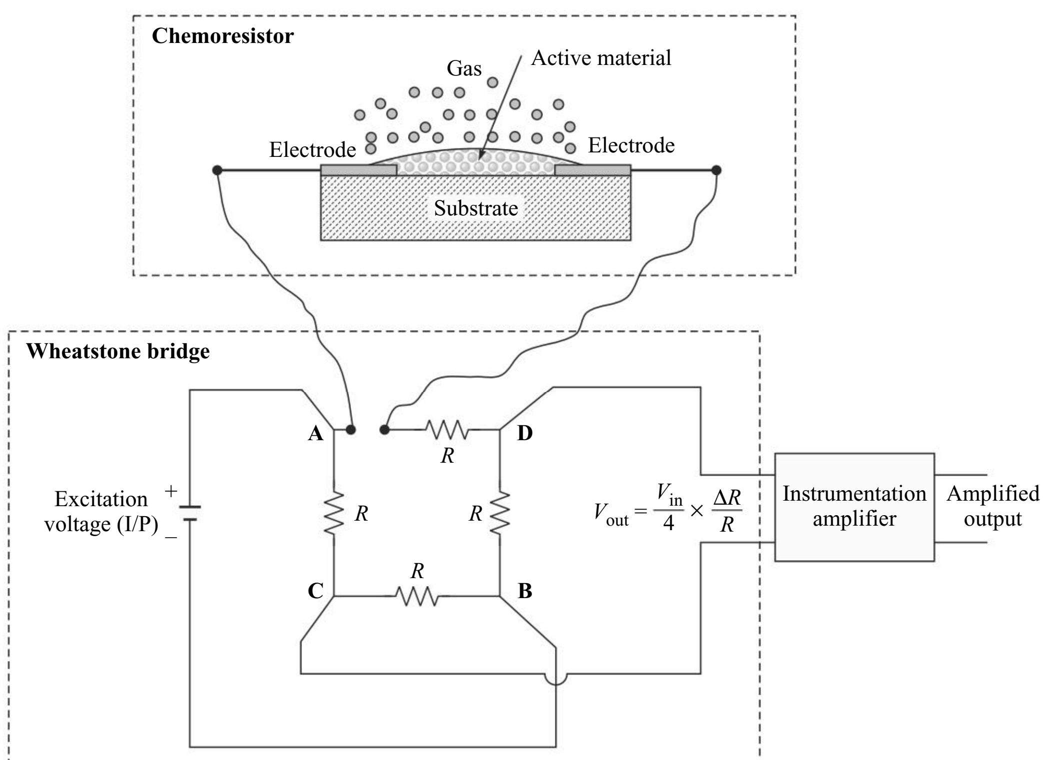


Fig. 11.3 *Schematic diagram of a chemoresistor*

selectivity can produce higher output level compared to a material with low selectivity. Reversibility is the ability of an interaction to return to its pre-interaction state. It means that initial conditions must be obtained when the input reaches initial values. When the sensors are non-reversible, we can distinguish between re-generable sensors and disposable sensors. In case of former one the initial conditions are restored through an additional chemical process and the latter type can only be used one time (medical sensors)

Many chemoresistors consist of resistor array formed from different polymers where only the resistance of a particular polymers change significantly depending on the gas to which it is exposed. Such type of configuration is useful in order to detect the specific gas and the method is known as pattern recognition technique. A typical pattern of varied resistance outputs corresponds to a particular gas. Experimental studies have revealed a number of factors that affect the sensitivity of the polymer-based chemoresistor. They are:

- The nature of the polymer
- The thickness of the film
- Electrode geometry
- Contact resistance
- Stability in the initial resistance (R_0)
- Film temperature
- Film wear and aging
- Hysteresis

Commercially available polymer-based chemosensors are used in the process control of alcohols, coffee, methane, ammonia, benzene, carbon dioxide, hydrazine and toluene.

11.7 CHEMOCAPACITORS

The chemocapacitor is a two terminal device. The principle of operation of chemocapacitor is similar to normal capacitor (Fig. 11.4). However, here the important point is the accountability of sensitivity. The chemocapacitors are fabricated from a diverse types of polymers such as polydimethylsiloxane (PDMS) and modified PDMS, polyetherurethane (PEUT) and polyepichlorohydrin (PECH). The electrodes of the capacitor are deposited on the top of an inert substrate. A sensitive polymer film is deposited over the electrode. This follows the deposition of the counter electrode layer. The change in capacitance is a measure of gas detection. The change in the capacitance due to absorption of analyte is related to three different activities:

- Adsorption on the polymer surface, giving rise of a new thin layer on the top of the polymer
- Absorption into the polymer phase in terms of changing the dielectric constant of the polymer
- Swelling of the polymer layer

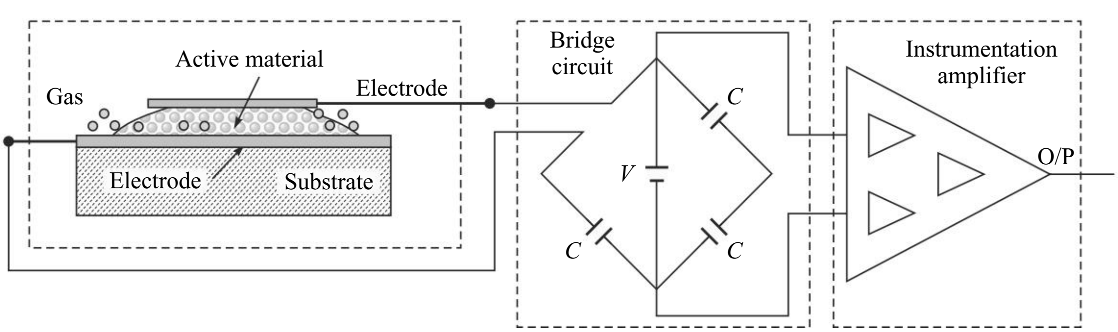


Fig. 11.4 *Schematic diagram of a chemocapacitor*

The chemocapacitor electrical connection is similar to chemoresistor as shown in the figure. In this case the change in capacitance affects the pre-balanced bridge circuit.

11.8 CHEMOTRANSISTORS

Environment pollution has become one of the major problems these days. There has been continuous technological development with regard to detection of toxins and related materials. Ion sensitive chemical sensors are useful in order to detect particular species in surroundings. Basic principle of operation of these sensors lies with the well-known Field Effect Transistors (FET). The FET based ion selective chemical sensors are called ISFET. Other name of ISFET is CHEMFETs (Chemically Modified FET). MOSFETs (Metal Oxide FET) based ISFET are common. The principle of operation of FET and MOSFET are necessary in order to understand the ISFET.

FETs are three terminal semiconductor devices and are of two types, *p-channel* and *n-channel* FETs. In each case, a semiconductor bar called channel of one type of semiconductor material is located inside a bulk of material of the other kind. That is, a *p*-type is inside an *n*-type, or vice versa. A pair of metallic contacts is placed at each end of the channel. If a voltage between these terminals is applied, a current can flow along the channel from one end to the other. The terminal, which produces charge carriers into the channel, is called the *source*. The terminal that attracts them is called the *drain*. A contact is made of the other type of material. This contact terminal is called a *gate* terminal. In an *n*-channel type device, the current carriers free to move along the channel are electrons, whereas in a *p*-channel type device the carriers are holes.

Figure 11.5 shows an *n*-channel FET with DC bias voltages V_{DS}, drain-to-source viltage and V_{GS}, gate-to-source voltage, appropriately applied across the terminals. Note the polarity of the connection. Much like a simple diode, the depletion region grows as the reverse bias across the PN junction is

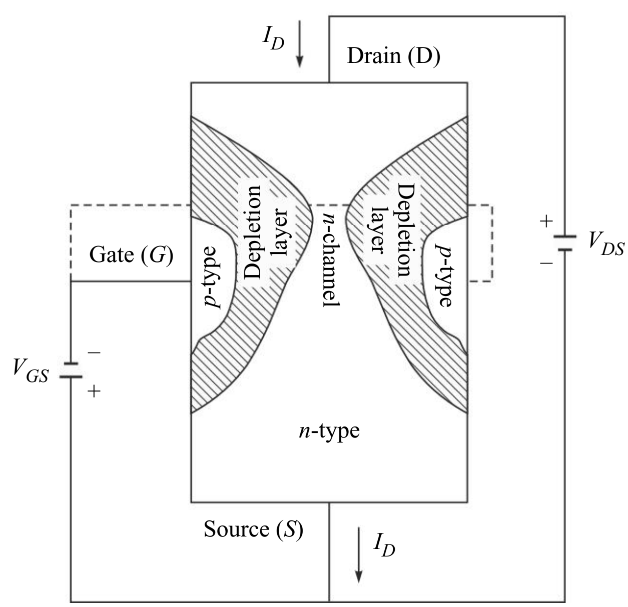


Fig. 11.5 *An n-channel JFET*

increased, thereby reducing the cross section of the conducting n-channel. Greater the reverse voltage, narrower is the cross section, the conduction path. Therefore, the reverse-biased gate voltage can control the flow of carriers through the channel. *A metal oxide semiconductor field effect transistor,* which is abbreviated MOSFET, is similar to the FET but has larger input impedance due to the existence of a thin layer of silicon dioxide in between the gate and the semiconductor channel. It is the applied voltages which determine which *n*-type region (in *n*-channel FET) provides the electrons and becomes the source, while the other *n*-type region collects the electrons and becomes the drain.

Because of the insulating layer next to the gate, input resistance of a MOSFET becomes greater than 10^{12} Ohms. MOSFET can deplete (make narrow) as well as enhance the channel with respect to resting state. Resting state is a state when the gate voltage is zero, i.e., $V_{GS} = 0$. Accordingly, we have two types of MOSFETs, one is called depletion MOSFET and the other enhancement MOSFET. This implies that the construction of MOSFET devices differs based on the channel size in the resting state. A depletion type MOSFET is called *normally on* MOSFET and the enhancement MOSFET called *normally off* MOSFET. If reverse voltage is applied in the gate junction of the former type, the channel becomes narrower with respect to resting width. On the other hand, applying a forward bias across the gate and source terminals creates the channel of the enhancement type. Note that at resting state current can flow in a depletion device but current never flows in an enhancement device due to the reason that there is no channel at all. Increasing the forward bias from the resting state forms the channel, in enhancement devices.

We will only consider *n*-channel MOSFET. Analogously the *p*-type MOSFETs will be understood. Figure 11.6(a) is in depletion mode and (b) is in enhancement mode. n^+ is referred to as heavily doped *n*-type materials. The substrate is *p*-type. Consider the depletion type device. When a negative voltage is applied, positive charges will be induced in the diffused *n*-channel effectively making the channel less wider and hence making the channel less conductive. Next, consider the enhancement type. When a positive voltage is applied across the gate, the minority carriers present in the *p*-type substrate form an inversion layer. Beneath this, the positive majority carriers will form an induced channel. The induced channel will cause the current to flow if a potential difference is established between the source and drain. Although, Fig. 11.6(b) describes a *p*-substrate based MOSFET, in practice the *n*-substrate based designs are preferred in order to make gate terminal forward biased with negative voltage. In such case the heavily doped materials will be p^+ instead.

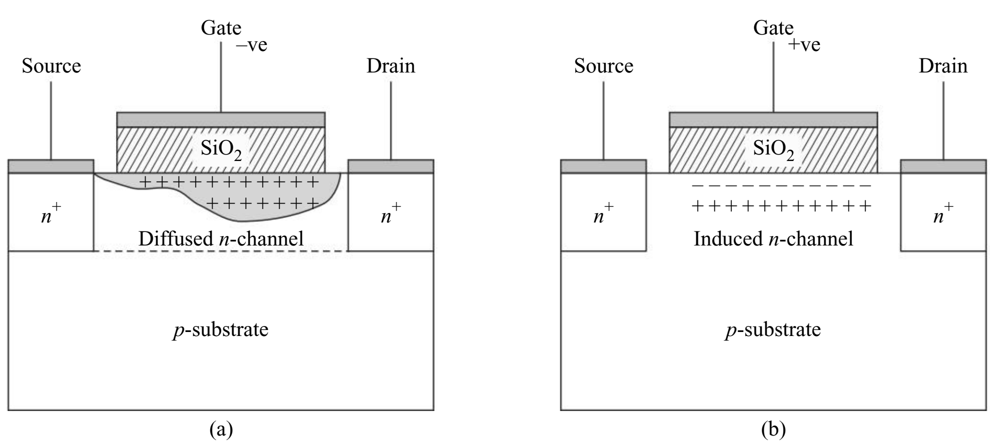


Fig. 11.6 *(a) Is in depletion mode, and (b) Is in enhancement mode*

In essence, the ion selective transistors are nothing but solid-state miniature device based on silicon technology. The construction of the MOSFET can be slightly modified in order to design an ISFET, which can be used as a chemical sensor. Figures 11.7 and 11.8 show the schematic diagrams of ISFET. The gate of the MOSFET is made of a chemical sensitive metal. The absorption of the chemical species by the metal changes the work function, thereby changing threshold voltage of the transistor.

Unlike MOSFET, the gate of an ISFET is a complex structure consisting of three components, namely (i) electrode, called reference electrode, (ii) a cavity-like space for analytes, and (iii) gate dielectric. The gate dielectric (Usually SiO_2) is coated with ion-selective material together constitute as the gate membrane. The ion-selective material is called electroactive material. ISFET is used for the detection of many types of ions, however, in each case the electroactive material should be different. For example, poly vinyl chloride containing valinomycin is an ion-selective electroactive material and primarily used for the detection of potassium ions. Similarly, for the measurement of nitrite concentration in agricultural and environmental pollution, the sensing layer is taken as PVC plasticized membrane, which contains uranyl salophen. The transistor can also be used as a pH sensor. In order to improve pH responsiveness silicon nitride (Si_3N_4), overlying the dielectric material (SiO_2) is used.

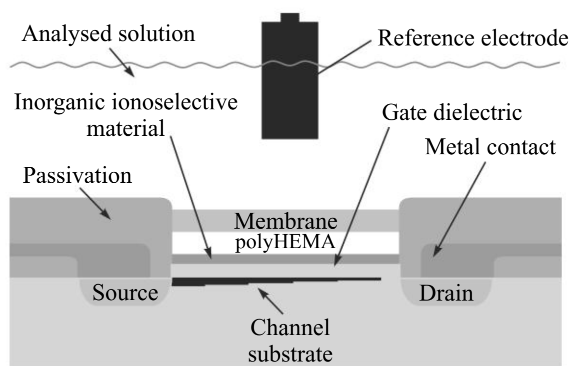


Fig. 11.7 *ISFET components*

The solution along with the membrane is considered as additional layers to the gate dielectric. These additional layers influence the threshold voltage of the ISFET developing a potential drop (called threshold voltage drop) in the gate to source loop. In other words, the ion concentration in the solution influences the gate potential, which in turn modifies the transistor threshold voltage. The threshold voltage is considered as an undesired limitation on the dynamic range of the transistor. The voltage range can be up to 2 volts. Mathematically, the threshold voltage of the ISFET can be written as:

$$V_T = \varphi_{ms} + 2\varphi_f - \frac{Q_{sc} + Q_{ss}}{C_{ox}} \tag{11.3}$$

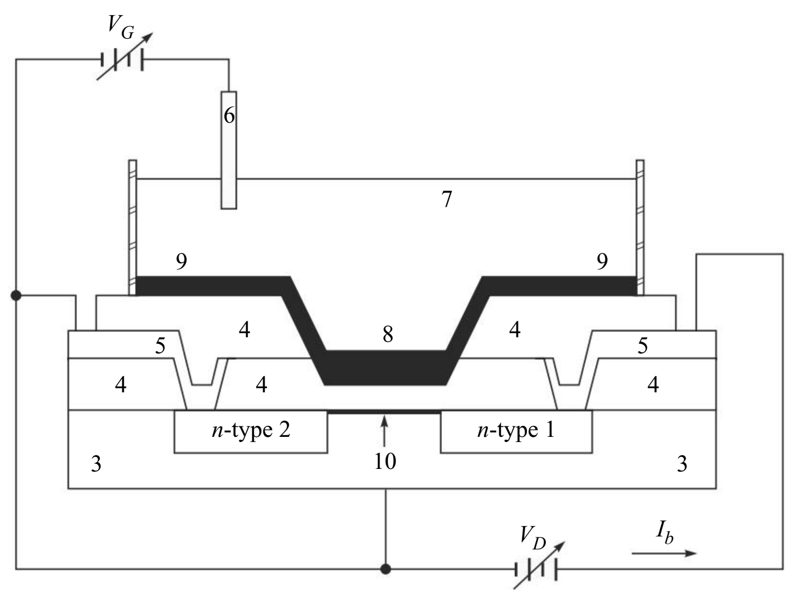
where V_T is the threshold voltage of MOS-based ISFET, φ_{ms} is the difference between the work functions of reference electrode and the gate dielectric, φ_f is called the Fermi potential of the semiconductor material used, Q_{ss} is the fixed surface-state charge density per unit area at the dielectric-semiconductor interface, Q_{sc} is the semiconductor surface depletion region charge per unit area and C_{ox} is the gate insulator capacitance per unit area.

Two methods are employed in order to detect this modification (i.e. the threshold voltage). If the applied gate bias potential is maintained constant then the influences at the interface of solution and

electroactive material are reflected through the changes of the transistor drain current, I_D. In second method, if the drain current is kept at a constant value then the change in threshold potential in the gate loop due to activity of the ion can be reflected in terms of an electrical voltage. In the latter case one can note that the transistor is always operated under the constant drain current mode, and the change in drain current due to the variation of the ion activity is compensated by adjusting the gate voltage. The amount of adjustment is the measure of ion concentration. The arrangement shown in Fig. 11.8 is simply an electrochemical sensor.

The sensitivity of the ISFET is expressed in gate voltage change per decade of the ion concentration. As an example the PVC plasticized membrane containing uranyl salophen receptor show linear response in the range 1–4 of pNO_2^- with a slope of 37 mV decade^{-1}.

The ISFET based chemical sensors are effectively attractive compared to other conventional types of sensors as they have good analytical performance, small size and low price. Small sized sensors require less reagent, resulting in faster, more accurate, less expensive operation.



1. Drain
2. Source
3. Substrate
4. Insulator
5. Metal contacts
6. Reference electrode
7. Solution (analytes)
8. Electroactive membrane
9. Encapsulant
10. Conducting channel
11. V_G = Gate voltage
12. V_D = Drain voltage

Fig. 11.8 *Schematic diagram of an ISFET*

Pollution or contaminations can be detected by optical detection methods, called spectroscopic techniques. But in reality some of the analytes are not optically active. Therefore, opto method may fail in those applications. In that respect electrochemical sensors take advantages. Further, the instrumentation part, i.e. data acquisition, amplification and calibration section in case of electrochemical sensors can be made compact and compatible. All these chemical sensors are either

based on chemically modified electrode sensors, microelectrodes, voltametric or potentiometric methods that can analyze many types of inorganic ions. The anions (negatively charged molecule) that are not electrochemically oxidizable can be analyzed using polymer modified electrodes. Moreover, because of miniaturization, more number of microelectrodes can be fabricated for simultaneous detection and monitoring of different anions. For example, multiple-element microelectrode array electrochemical sensors can monitor Cu^{+2}, Pb^{+2}, CN^{-}, Ag^{+1}, and pH simultaneously. These are called multiple-element microelectrode array electrochemical sensors. Materials such as gold, platinum, carbon or silver are commonly used for electrode. These raw individual electrode elements are modified chemically selective by utilizing chemical or physical techniques.

Example 11.1 pH Measurement along with ISFET Specifications: (Source: CSIC, Instituto de Microelectrónica de Barcelona)

General specification

Substrate:	p-type 4-inch silicon wafers
Chip dimensions:	2.5 × 2.5 mm
Gate length:	10~12 μm
Gate width:	≥ 400 μm
Gate structure:	Silicon oxide/Nitride
Devices per chip:	1 ISFET and 1 MOSFET
Position of the ISFET gate on the chip:	A gap between ISFET gate and chip edge and/or metallisation is at least 500 μm wide

Electrical parameters

Operational drain voltage:	$V_d = 0.5 - 2.0$ V
Operational drain current:	$I_d = 0.1 - 1.0$ mA
Transconductance:	$g_m > 0.5$ mA / V
Threshold voltage:	$-2 <$ Vth< 0.7 V at pH 7
Leakage currents:	$I_l < 0.1$ nA

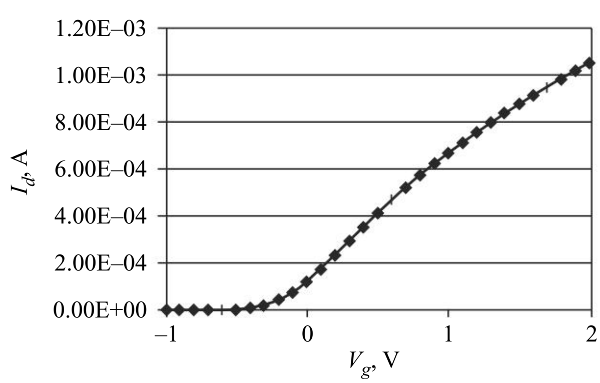


Fig. E-1 *Current–Voltage characteristic of pH-ISFET devices measured at VD = 0.5 V (Courtesy:* Centro Nacional de Microelectrónica, IMB*)*

Chemical parameters	
pH sensitivity, *S*	50–58 mV / pH (20–25°C)
Linear range	1–13 pH
Precision	0.05 pH
Selectivity coefficients	< 10–5 to alkaline metal ions
Long term drift	< ± 0.5 mV/h
Working lifetime	> 6 months

11.9 ELECTRONIC NOSE (E-NOSE)

Electronic noses (e-nose) are sensors, which mimic the sense of smell. They are in fact gas sensors. The difference between the other types of chemo-/biosensors and e-nose is that the design and construction of the latter is more complex, as they possess more number of elements and circuitry. The e-noses are integrated with signal processing building block for easy and effective mapping of various odors. There are lots of applications of electronic nose in many sectors. Some of them are listed as follows.

- Industrial processes—chemical sensing
- Food processing industry—food quality control
- Cosmetic industries—perfume, fragrance sensing, etc.
- Environmental bioprocess—pollutants, toxins detection and air quality measurement and analysis
- Medical and health science—medicine, body functions, microbial detection

E-nose systems use thin or thick film semiconducting (inorganic and organic) materials, such as metal oxide chemoresistors or similar devices. The principle behind the e-nose is based on the fact that the absorption of gases modifies conductivity of a sensing membrane, which is simply metal oxide semiconductors (MOS), or conducting polymers (CP) called the active material (Fig. 11.9). Table 11.2 shows the type of sensor that uses the active material for specific target odors. There are two types of e-nose what are being called as chemosensor and biosensor. In many situations, the chemosensor uses the synthetic material and the biosensor uses the biomaterial.

Table 11.2 *Sensors used by the electronic nose group to detect odors*

Active material	*Sensor type*	*Target odours/gases*
Metal oxide thick films	Chemoresistors, e.g. commercial TGS & FiS.	Alcohols, ketones, combustible materials.
Metal oxide thin films	Low-power chemoresistors. as above.	NO_x, H_2, NH_3.
Conducting polymers	Chemoresistors, chemodiodes.	Polar molecules, organic vapours, Phthalocyanines
Electrochemical cells.	Electrodes	NH_3, CO, ethanol, etc.
Lipid layer	Chemoresistor.	Organic vapors.

The e-nose is a system that incorporates arrays of metal oxide and conducting polymer odor sensors in which many analytes are determined simultaneously. Figure 11.9 depicts a single unit of an e-nose. An e-nose has several such units and corresponding signal processing circuitries that are interfaced to a

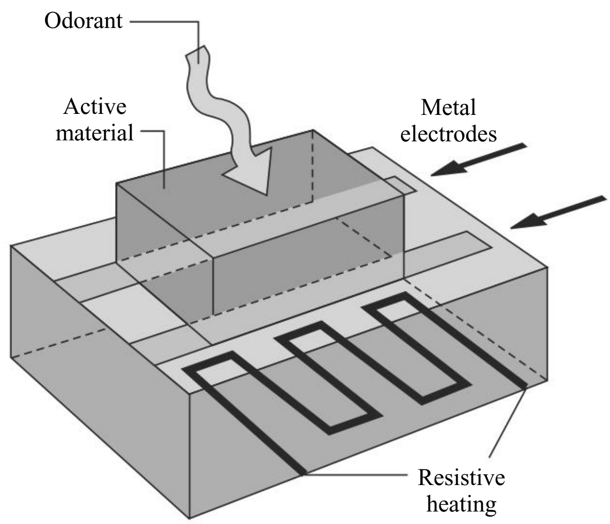


Fig. 11.9 *A unit in a typical e-nose*

processing unit incorporating advanced data processing algorithm. The arrayed e-nose system is a pattern-recognition system capable of recognizing simple or complex odors. The sensor system is composed of two integrated sections receptor or transducer, which is simply a highly selective recognition material and an analytical-producer section that analyses, maps and produces the odor class. The latter section utilizes the advanced data processing techniques such as neural networks and fuzzy logic characterizing an intelligent e-nose.

There are mainly three ways of implementing the sensor arrays depending upon the application types. They are

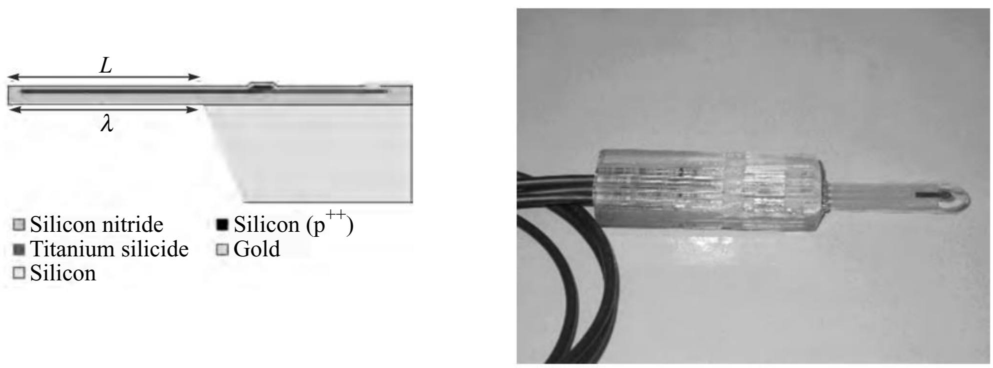
- Redundant-sensor array systems
- Selective-sensor array systems
- Cross-selective-sensor array systems

Identical chemical sensors form the redundant-sensor array systems. The selective-sensor array systems are a group of sensors with different selectivity. Each sensor can respond to a single odor component in the multi-component mixture. The third type system integrates a group of sensors with a cross-selective response pattern. Rather than measuring a number of single parameters separately, they can measure complex phenomena such as quality or process state.

11.10 DNA SENSORS

DNA (Deoxyribonucleic Acid) is a molecule in the living cell. The building blocks of DNA molecule consist of two long chains of amino acids. The chains usually twist around each other. The living being uses DNA to store its genetic code that describes everything about the organism. DNA carries the whole genetic information of all living creatures. In a particular organism, each and every cell possesses identical copies of the DNA. This implies that when a cell divides in normal growth both the cells contain the copy of the DNA of the previous one. A cancer is a growth of cells whose DNA has become distorted. The sectional parts of a DNA molecule are called genes. Genes determine how to build something genetically.

Genomic information is revolutionizing the life sciences, in particular health care. Specific fractions of the human DNA reveal valuable data. These data can be utilized for health care and medication purposes. The detection of genomic information is a remedial measure which can be performed at the early stage of any dreadful diseases. The innovative tools for genetic analysis for biomedical needs are still under development. One such development is the eSensor™ DNA Detection System by Motorola Inc. using which reliable DNA testing can be performed. The basic element of the DNA analysis system is a DNA probe(s) (Fig. 11.10(a)). The DNA analysis system with probe is called DNA sensor (Fig. 11.10(b)).



(a) (b)

Fig. 11.10 *(a) A packaged DNA probe, (b) A DNA probe showing the material from which it has been faricated*

Some important genomic analysis applications are listed below.

- Forensics
- Genetic screening
- Stress-response analysis
- Antibodies gene expression in transgenic cells
- Biowarfare agents
- Pathogens like throat bacteria
- Drug discovery

The development of DNA sensors is considered as the most innovative molecular biology technology. These micro or nano systems allow an easy, accurate, fast and reliable way to analyze many DNAs simultaneously. The technology is based on the recognition of oligonucleotides[2] by their complementary genomic sequences[3]. The process of recognition is called hybridization. The hybridization reveals the presence of small amount of a specific segment of DNA by changing some characteristics of the DNA probe. The hybridization can be achieved in three ways.

- Optical method
- Surface stress
- Electrical detection

11.10.1 Optical Method

Optical method is a traditional method, in which the color of the solution changes. University of Rochester researchers have designed a simple and inexpensive sensor that can detect specific

[2] An oligonucleotide corresponds to a very short DNA single strand

[3] The relative order of base pairs, whether in a DNA fragment, gene, chromosome, or an entire genome.

sequences of DNA. The sensor contains hairpin-like stalks of DNA that straighten to expose fluorescent molecules attached to their ends when they combine with a given sequence of DNA. A minute drop of solution containing the DNA to be identified is placed onto the sensor. The solution changes its color.

11.10.2 Surface Stress Method

Surface stress based hybridization process uses a micro cantilever structure (Fig. 11.11(a–c)). The cantilever is coated with the detector film that reacts with the biomolecules of interest. It is required to prepare the cantilever before it can work as a DNA probe. In the presence of DNA molecules biochemical reaction takes place. A biochemical reaction at the cantilever surface changes the surface stress. The surface stress causes a bending of the cantilever. A piezoresistor integrated with the cantilever helps to measure the bending variable (e.g. curvature), which indeed is the measure of the presence of DNA. The change in surface stress is transformed through piezoresistor for appropriate instrumentation. Piezoresistive material has the property that it changes its resistance when strained.

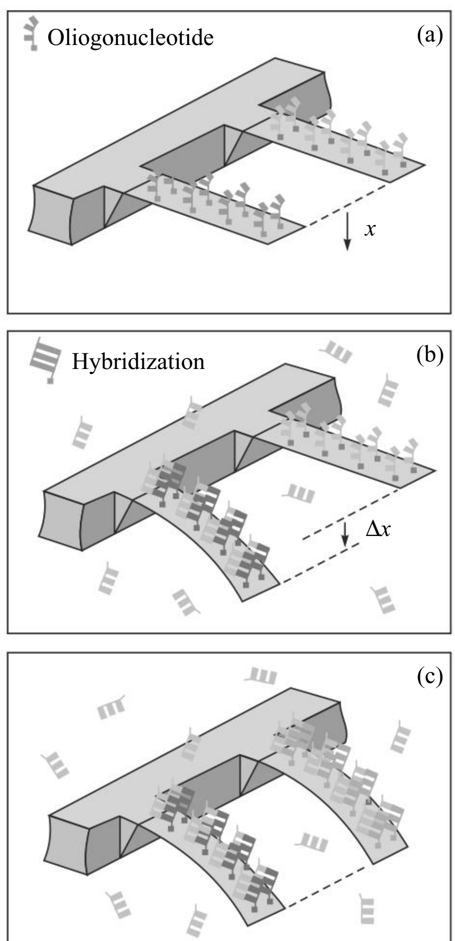


Fig. 11.11(a–c) *Surface stress hybridization*

Most materials change the cross section and length under a load and change their resistance values. However, the resistance change in case of piezoresistive material is far larger than that which can be accounted for by geometrical changes. Wheatstone bridge can be integrated to measure the change in resistance caused by surface stress. When voltage is applied to Wheatstone bridge with the resistors of identical resistance R, the differential output voltage can be measured by using the following formula:

$$\Delta V_{\text{out}} = \frac{1}{4} \times \frac{\Delta R}{R_{\text{ini}}} V_{\text{in}} \tag{11.4}$$

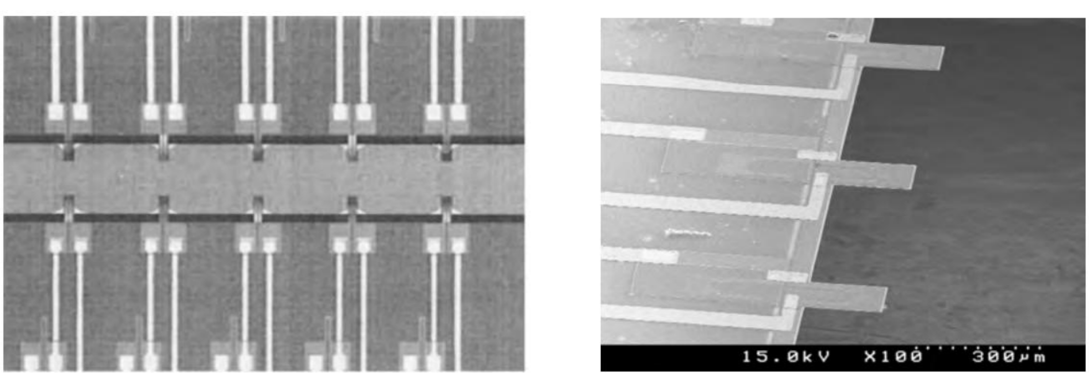
A complete description of the above equation is presented in Chapter 4. A very small amount of bending can be measured by this method. Further, many cantilevers can be used simultaneously with different detector strand in order to detect different biomolecules simultaneously. Such arrangement is called cantilever arrays. Up to 20 or more numbers of cantilevers can be fabricated on the same platform. Figure 11.12 shows SEM (Scanning Electron Microscopy) photographs of a piezoresistive based cantilever arrays. The cantilevers are designed on two sides of a channel (Fig. 11.12(a)) through which the analytes are transported. Such a device can be understood as a CLOC device. The size of the cantilever is about 120 μm × 350 μm. The stress-based hybridization is ideal where limited sample preparation is expected.

Another promising method of detecting surface stress in a cantilever is known as static deformation approach. Static deformation approach measures the radius of curvature of the cantilever bend. The expressions for radius of curvature and the displacement of the cantilever tips are given as follows.

$$R = \frac{Et^2}{6(1-\nu)\,\Delta\sigma} \tag{11.5}$$

$$\Delta d = \frac{3l^2(1-\nu)}{Et^2}\,\Delta\sigma \tag{11.6}$$

where, R is the radius of curvature of the cantilever, ν is Poisson's ratio, E is Young's modulus, t is the thickness of the cantilever, $\Delta\sigma$ is the differential surface stress, l is the length and Δd is tip deflection (displacement).



(a) (b)

Fig. 11.12 *SEM photograph of a piezoresistive based cantilever array (a) 20 cantilevers, 10 in each side of a microchannel through which sample can flow; (b) Close view of another SEM photograph*

11.10.3 Electrical Detection

The electrical characteristic based hybridization process uses microelectrodes, which are made up of gold. DNA probes are usually attached to the electrodes (shown in Fig. 11.13). Similar to cantilever based method, the probe for each electrode must be prepared beforehand. In essence, the DNA has to be coated onto the probes prior to their operation. Each gold electrode integrated with active probe now has a specific DNA strand with known molecular sequence. The pairs of electrodes are placed in a semicircular compartment. A solution of unknown sequence, i.e. the target DNA strand is then brought in contact with the probes. Compatible DNA strands bind to the attached probes and the unreacted strands do not. Any DNA that hybridizes with the probe is the DNA to be detected (Fig. 11.13(a) & (b)).

In resting state the electrical resistance between the two opposite electrodes is infinite. Upon hybridization, a bridge like structure is formed on the electrodes. Because of the formation of the bridge, the electrical resistance between the electrodes drops from infinite to few thousand Ohms. The change in resistance value is the measure of the presence of DNA. Appropriate instrumentation can be integrated to observe the effect. Figure 11.13(c) shows how a typical bridge has been formed by the use of an electrical characteristic based hybridization process using microelectrodes attached with DNA probe.

One can note that the electrical characteristic based detection of a specific DNA molecule fragment is very easy and sensitive compared to other methods. In particular, the electrical detection method

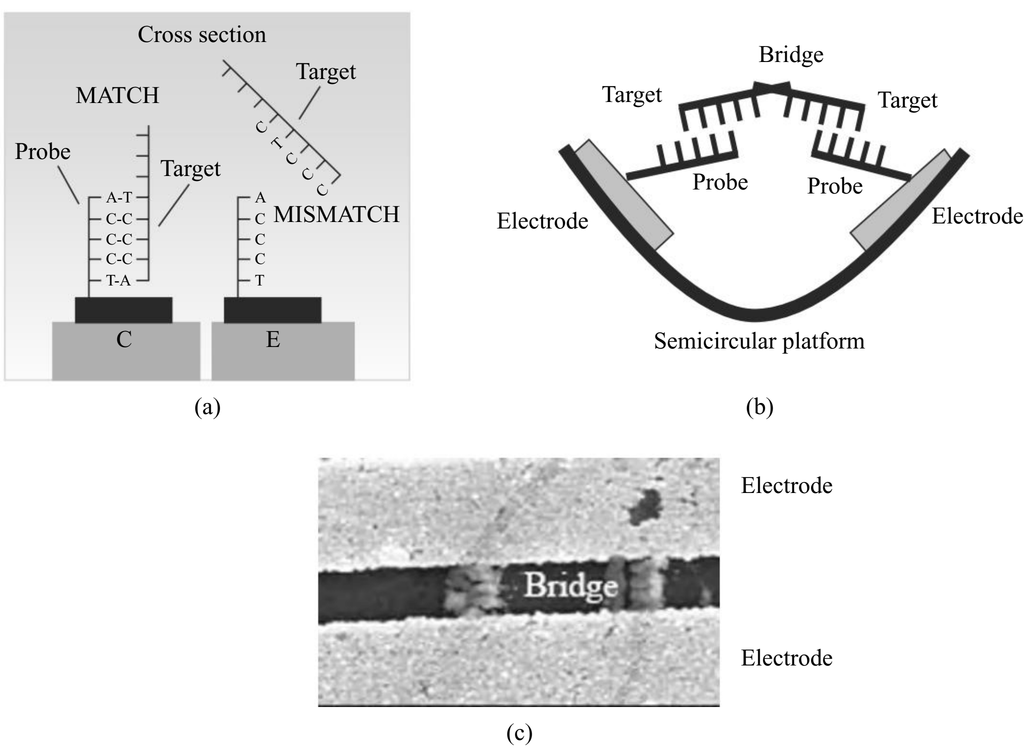
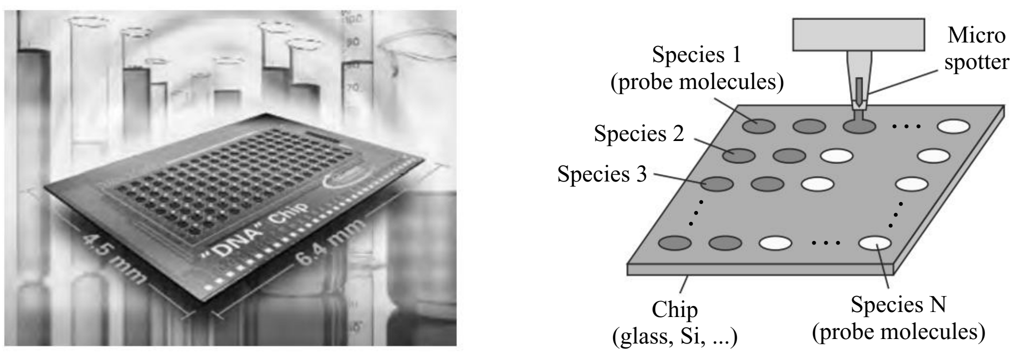


Fig. 11.13 *(a): A simple example of matching of DNA strands, (b) More practical arrangement of electrodes, matching of DNA strands and formation of bridge, (c) The detection of a single specific DNA molecule fragment by electrical hybridization*

provides several advantages including the fact that it avoids expensive and complicated optical set-up and nonlinearity present in the stress measurement and claimed to be robust as far as operation is concerned. Further, it is easier to design custom based DNA probes in order to detect specific DNA strand of known sequence, which could be the bacteria, viruses or even HIV DNA.

11.10.4 DNA Chip

Fully integrated electronic DNA sensor arrays on CMOS are available in the form of DNA chip. Figure 11.14 shows a DNA chip. The chip contains arrays of predefined spots. An array of 16×8 is typical. On these spots DNA probes are prepared. On each position different probes are developed. DNA probes are prepared prior to their operation. Single-stranded DNA molecules known as DNA sequences are deliberately spotted on the surface of the chip at the predefined positions. The spotting is carried out by a DNA microspotter as shown in Fig. 11.14(b). The process of development of probe is also called immobilization. The analyte containing target molecules to be detected is applied to the spots. Hybridization occurs in case of matching DNA strands. The signal is taken out by utilizing surface stress or electrical property principle as described above. Through the spots the DNA conformant electrical signals can be read out. An arrayed sensor has electrodes built into it with spacing approximately 1 μm arranged within a circular compartment with typical diameter 100–300 μm.



(a) (b)

Fig. 11.14 *(a) A DNA chip; (b) Preparation of DNA probe using a microspotter (Source: Research trends,* A publication by Corporate Research, October 2002*)*

11.11 MASS SENSITIVE CHEMOSENSORS

Electrochemical chemosensor produces output based on its involvement in chemical reaction, which in turn changes electrical properties such as resistance or capacitance. There exist another group of chemosensors, which are sensitive to physical properties such as mass and shape of the solution (Not the mass and shape of the active material). These types of chemosensors are called *mass-sensitive* chemosensors or *physical* chemosensors. The problems encountered by reaction-based chemosensors can be circumvented by the mass-sensitive chemosensors. Mostly, mass-sensitive sensory system utilizes the principle of *spectroscopy*. In many applications the former (reaction-based) types of chemosensors show poor repeatability (described in Chapter 14) because of the following reasons.

- Selection of a reversible, reproducible and stable chemical active material
- Fabrication and interfacing of dielectric layer and active material layer

- Chemical sensing in fluids often exhibits clogging and contamination of the transducer (active material)

11.11.1 Spectroscopy and Mass Spectroscopy

The study of spectral lines from different atoms and molecules is known as spectroscopy. Spectroscopy is a reliable, attractive and promising method of studying chemicals. Spectroscopic instruments have long been used as macro level instruments. Recently, however, micro spectroscopes have been developed. The principle of operation of both types of spectroscopes (macro and micro) is same. Electromagnetic (EM) radiation at some frequency is allowed to *interact* with the chemical of interest. Then the properties of that radiation, such as amount diffracted, emitted, scattered and absorbed is monitored. The frequency of the EM radiation is determined by the energy levels associated with the property of the target chemical sample. Although, the theory of spectroscopy is simple to understand, the spectroscopy instrument is considered as complex, as there are many subsystems and the interfacing amongst them is crucial.

The interaction between the chemical and the EM wave can be achieved in many ways. Mostly, the EM radiation interacts with the chemical in its ionized form. Note that ions are easier to handle and manipulate than neutral molecules. Study of analytes (chemical or biomolecule) in their ionised state with the interaction of EM signal is actually known as *mass spectroscopy*. The instrument that carries mass spectroscopy is called mass spectrometer (MS).

11.11.2 Principle

The chemical to be detected is first ionized. The ionized chemical is then passed through a magnetic field of pre-defined direction. The ions are accelerated and collected at the detector. The particles travel along a curved path on a radius that is proportional to the mass to charge ratio (m/z). The curvature will vary for different chemicals, since the ratio varies from chemical to chemical. The MS consist of three distinct regions: ionizer, ion analyzer and detector and display as shown in Fig. 11.15(a).

11.11.3 Ionizer

One important method of ionization of analytes is achieved through a process known as electrospray (Fig. 11.15(b)). The SEM photograph of an electrospray is shown in Fig. 11.14(e). Besides electrospray, the two other methods are electron impact-based method and chemical ionization-based method. In the former case the sample is imparted with electrons, whereas in the latter case the ionization is achieved through chemical reaction.

The applied field directs the released ionized molecules. The electric field also changes the shape of the ionized droplet from spherical to conical. The cone shaped ionized droplets are called Taylor cone. Schematic of structure of a microcapillary tube interfacing the analyte reservoir and nebulizing gas reservoir is illustrated in Fig. 11.15(c). The sprayed ions are actually passed through a section called MS inlet (see Fig. 11.15(d)). MS inlet has three components: counter electrode, sampling orifice and skimmer. The purpose of counter electrode is to accelerate the ions. The sampling orifice and skimmer do not allow the ions that get deviated away from the line of flight to pass through towards the detector.

Electrospray method uses silicon or polymer based capillary tubes that can be integrated with the MS inlet. The capillary tube has electrospray nozzles. The tip of the nozzle is usually made from silicon nitride. Preparations of silica nozzles require complex procedures and also have limitations with respect to yield stress for which polymer-based microcapillary nozzles are drawing interest. Parylene-C is

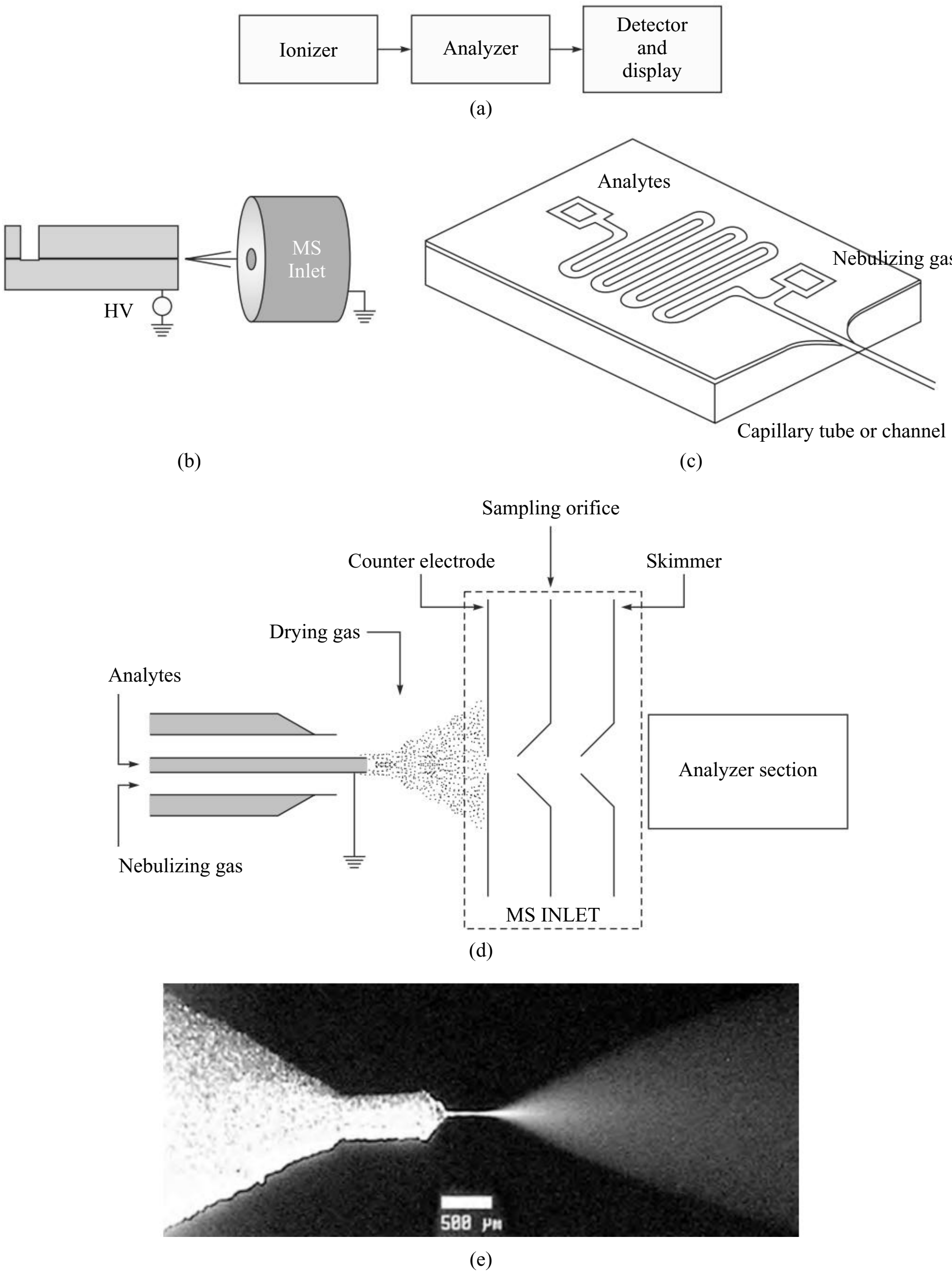


Fig. 11.15 *(a) Mass spectrometer building blocks, (b) A typical micronozzle interfaced with MS inlet, (c) Schematic of structure of a microcapillary tube interfacing the analyte reservoir and nebulizing gas reservoir, (d) Detail schematic diagram of micronozzle and MS inlet section, (e) SEM of a micronozzle in operation, (Courtesy (b &c):* California Institute of Technology and Beckman Research Institute; *(e)* Dahlin et al. *Analyst*, 130/2, 2005*)*

biocompatible and the microchannel or microcapillary fabrication process is not complex. Moreover, due to its low tensile stress, the tube can be fabricated with high aspect ratio. Typical dimensions of a microcapillary nozzle can start from few micrometers to few millimeters. Microcapillary with length 2.5 mm and with a 5 μm × 10 μm orifice size at the tip is common.

11.11.4 Analyzer

The next region is called ion analyzer. Ion analyzer pumps away the uncharged and fragmented molecules, while allowing the ionized molecules to move towards the detector. There are mainly four ways of pumping away the unwanted ions. They are achieved by using quadrupole, sector field (magnetic and/or electrostatic), and time-of-flight (TOF).

Quadrupole: A quadrupole assembly consists of four parallel rod type poles through which the ions being separated are passed. The poles are supplied with fixed DC and an alternating radio frequency voltage. Depending on the produced electric field by the quadrupole, only ions of a particular m/z will be focused longitudinally and deflected toward the detector. Other ions will be attracted into the rods. The length and diameter of the rods determine the mass range and required resolution.

Magnetic: Sector field mass analyzer uses magnetic or electric fields to deflect the ions through a curved trajectory. The electric field strength and the RF frequency can be varied in order to tune a particular type of ions for the generation of mass spectrum. The magnetic field strength required to cause a deflection is given by,

$$B^2 = \frac{2mV}{zR^2} \quad \Rightarrow \quad \frac{m}{z} = \frac{R^2 B^2}{2V} \tag{11.7}$$

where, V is the applied voltage, m is the mass of the ion, z is the charge, R is the radius of curvature, B is magnetic field.

Time-of-flight: In TOF ion analyzer, to extract the ions an extraction voltage is applied that accelerates them into the field-free drift zone. The TOF ion analyzer is the simplest type of common mass analyzer with high sensitivity. The basic formula that be used for the design of TOF analyzer is given by the following equation.

$$\frac{m_i}{z_i} = 2eEl_s\left(\frac{t_i}{l_d}\right)^2 \tag{11.8}$$

where m_i = mass of analyte ion, z_i = charge on analyte ion, E = extraction field, t_i = time-of-flight of ion, l_s = length of the source, l_d = length of the field-free drift region and e = electronic charge (1.6022×10^{-19} Coulomb).

11.11.5 Detector and Display

With regard to detectors, there are many types. A common type that is in use works by producing an electronic signal when struck by an ion. The detector collects the ions according to m/z and electronically records the abundance of each m/z. Integration mechanisms are applied in terms of accumulating (summing) each recoded signal in order to report what is known as *mass spectra*, which is a measure of mass-charge ratio that strikes the detector over a period of specified time.

The most useful information revealed from a mass spectroscopy is the molecular weight of the sample, which is unique for a chemical sample. The mass spectra are measured in terms of relative

abundance. Figure 11.16 shows an example of how the mass spectra are represented. The ratio of peaks containing ^{79}Br and its isotope ^{81}Br confirms the presence of bromine in the compound.

The size of an integrated micro mass spectroscope (IMMS) bench is approximately 14 mm long contrast to 3 feet macro spectroscope. Although the IMMS bench is 14 mm long, the important parts such as capillary channel with nozzle, MS inlet and the detector are very small in structure, in the order of micrometer.

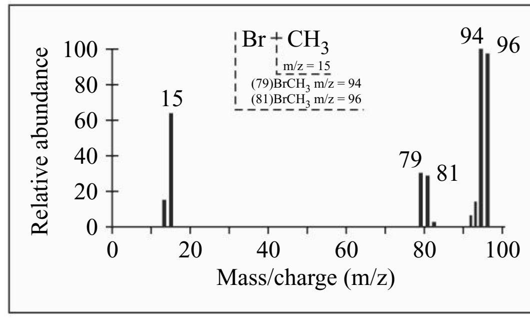


Fig. 11.16 *A typical mass spectra (Source: Department of Chemistry, U. of Arizona)*

11.12 FLUORESCENCE DETECTION

Another mass-sensitive device is based on fluorescence detection. Here a planer waveguide is used as the transducer. A material based medium that confines and guides the electromagnetic (EM) wave is known as waveguide. The EM wave is propagated along a path defined by the construction of the guide. A waveguide fabricated in a flat material such as thin film is known as planer waveguide. The detection principle exploits the inherent fluorescence properties of the analytes (chemical). The arrangement consists of three building blocks such as the irradiation source, a planner waveguide coated with sensitive polymer layer and a detector as shown in Fig. 11.17. The fluorescent analytes (chemical) are initially exposed to ultraviolet (UV) light or other high-energy ray (irradiation source) with wavelengths shorter than those of visible light. Because of exposure, the irradiation takes place. Irradiation is the process whereby the sample is exposed to radiation source. Irradiation has to take place rectangularly to the waveguide surface. The emitted fluorescence light is coupled into the waveguide material and directed toward the detector placed vertically. The fluorescence intensity is monitored rectangularly to the irradiation. By varying the size of waveguide surface high gain fluorescence emission is possible. The sensor response time can be improved by varying the thickness of the polymer layer.

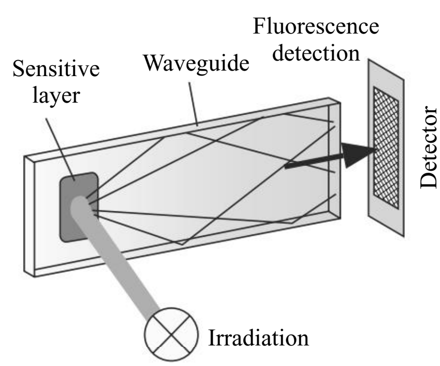


Fig. 11.17 *Mass sensitive typical fluorescence detection scheme (Source: J. of Royal Society of Chemistry,* 126, pp 766–771, 2001*)*

11.13 CALORIMETRIC SPECTROSCOPY

Recently a new invention known as calorimetric spectroscopy (CalSpec) has emerged. The CalSpec technology is originated in Oak Ridge National Laboratory (ORNL) and has received a US patent. The

technology is based on femtojoule sensitive thermal detector, which is coated with chemical layer. The coated chemical should be appropriately selected so that they are selective to the target chemical. The principle of operation of CalSpec follows. The CalSpec primarily consists of four building blocks such as (Refer to Fig. 11.18):

- A broadband light source
- A monochromator
- Thermal detector array
- The electronics read out circuitry

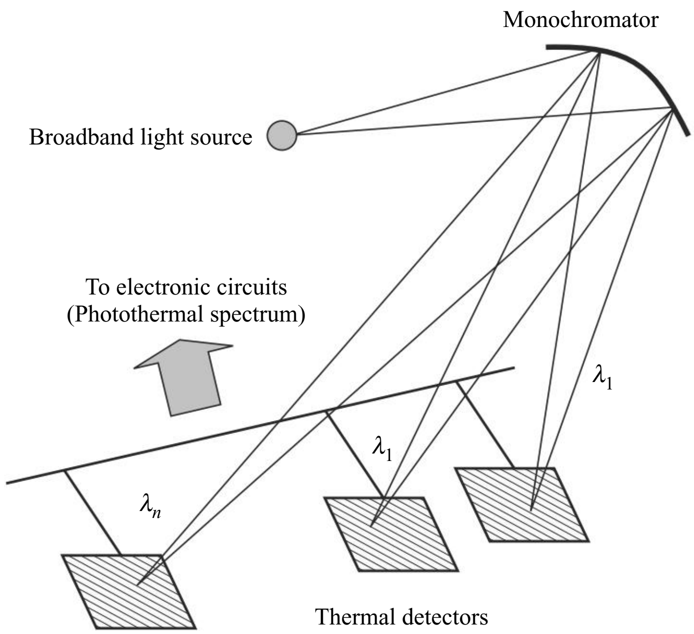


Fig. 11.18 *The concept of Calorimetric Spectroscopy (CalSpec)*

The broadband light source produces light signal containing a wide range of wavelengths. The thermal detector array detect the change in temperature. The thermal detectors are coated with sensitive material. When the molecules of the chemical under test come in contact with the thermal detectors, they are absorbed into the detector surface. The monochromator is integrated to produce narrow part of the broadband spectrum. In essence the role of the monochromator is to produce a single spectral line. It illuminates the chemically absorbed thermal detector. The temperature of the film changes upon irradiation. The temperature change of a particular thermal detector among the array, in effect, is proportional to the number of incident photons, which is in turn depends on the number of molecules absorbed on the detector surface. The measurable output is defined in terms of what they call *photothermal signature*. The read out circuitry is interfaced to process, calibrate and subsequently display photothermal signature. The display form of photothermal signature is called photothermal spectrum. The array of thermal detectors is very sensitive and therefore called femtojoule detectors. The detectors are exposed to different wavelengths. Pattern recognition technique can be employed to characterize the target chemical. The choice of number of detectors in an array determines the sensitivity of CalSpec. Further, the performance of this calorimetric spectroscopy is determined by the selectivity of the coating material.

11.14 SURFACE ACOUSTIC WAVE (SAW) SENSORS

Acoustic waves have been exploited for a wide range of applications including design of sensors for physical sensing, chemical sensing and biosensing. The acoustic wave principle requires knowledge on material properties, design constraints, and the sensing mechanisms. In this section only the sensing mechanism will be described. Other aspects can be found elsewhere in this book. However, it suffices to say that predominately quartz materials (piezoelectric) are used as wafer materials for the fabrication of such sensors. Some of the quartz materials are silicon dioxide (SiO_2), lithium tantalate ($LiTaO_3$), and lithium niobate ($LiNbO_3$). However, non-piezoelectric substrate such as silicon (Si) and gallium arsenide (GaAs) can also be used.

11.14.1 Principle

In principle, the surface acoustic wave (SAW) detection method uses a thin membrane, made of piezoelectric material vibrating at certain high frequency. The surface of the membrane is coated with selective material so that the chemicals to be detected get attracted towards the membrane. If present, those chemicals stick to the membrane, slowing its vibrations. This is called acoustic attenuation. The change in frequency is referred to as *resonant frequency shift* (RFS). The resonant frequency shift is given by,

$$\Delta f = k f_0^2 h \rho \tag{11.9}$$

where Δf is the frequency shift due to mass loading, k is a constant, f_0 is the resonant frequency, h is the height of the layer, ρ is the density of the layer. The principle can be further described as follows.

There are mainly five parameters that have direct influence on the response (resonant frequency shift) of such acoustic device. These are

- Conductivity
- Viscoelasticity
- Pressure
- Temperature
- Mass

The conductive thin film coating perturbs the electrical boundary conditions for the acoustic wave. Many materials possess solid-like (elasticity, strength, etc.) as well as liquid-like characteristics such as flow depending on time, temperature, rate and amount of loading. Materials that demonstrate both elastic and viscous behavior under applied stress are called viscoelastic material and the property is called viscoelasticity. The variation of viscoelasticity can be exploited. Acoustic devices are sensitive to mechanical loading, such as pressure. The wave is affected by the strain induced by the pressure. The effect of temperature induces dimensional dilatations. Mass loading refers to an influence, in which by the application of some chemical coating to the substrate of the device, the resonant frequency can be disturbed. *We are here, concerned with mass loading influence based surface acoustic wave (SAW) sensors.*

11.14.2 Types of SAW Sensors Based on Propagation Mode

The way the acoustic wave propagates through or on a piezoelectric substrate describes the acoustic devices. Primarily, their velocities and displacement directions distinguish the acoustic waves. In this respect, many combinations are possible, depending on the substrate material and boundary conditions. A wave propagating through the substrate is called a bulk wave. If the wave propagates on the surface of the substrate, it is known as a surface wave. Accordingly, we have bulk acoustic wave (BAW)

devices and surface acoustic wave (SAW) devices. We will continue discussing only the SAW devices. In 1887, Lord Rayleigh discovered the SAW mode of propagation. He found that the waves have a longitudinal and a vertical shear component that can couple with a medium that is in contact with the surface of a device. The coupling strongly affects the amplitude and velocity of the input wave. This feature enables SAW sensors to directly sense mass and mechanical properties.

11.14.3 Fabrication

Photolithography process is an acceptable method for the fabrication of interdigital transducer (IDT). First of all the wafer is polished and cleaned. Aluminium is mostly used for the IDT. Aluminium is uniformly deposited over the quartz substrate, followed by deposition of photoresist. As always, photoresist is spin-coated. The resist is then baked to harden using exposing UV light energy through a mask with opaque areas corresponding to the areas to be metalized on the final device. The exposed areas undergo a chemical alteration that allows them to be removed at a latter stage. Once exposing is over, the etching process is carried out to remove the selected portion of the aluminium material. Then the remaining photoresist is removed. The pattern of metal remaining on the device is the IDT. By changing the geometrical dimensions such as length, width, position, and thickness of the transducer the performance of the sensor can be optimized.

11.14.4 Notes on Boundary Conditions

Figure 11.19 shows the basic structure of such a device. The basic element (substrate) is quartz. They are also called mechanical resonators. The Q-factor of the resonator is very high. A resonator behaves

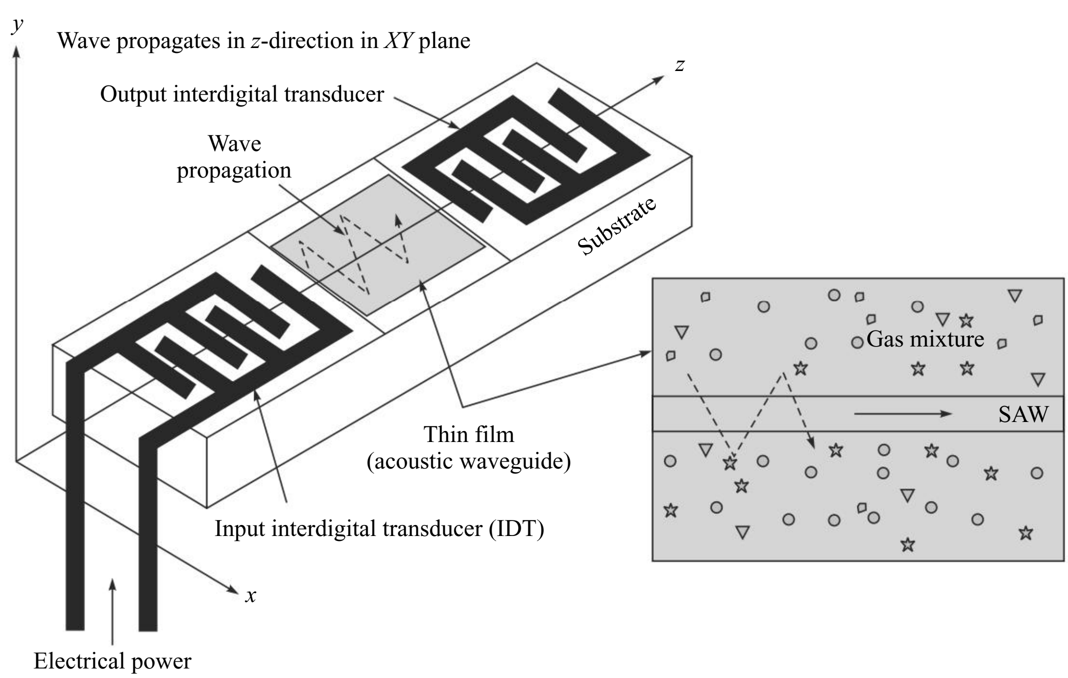


Fig. 11.19 *Basic structure of such device*

like a stress-free bulk. Because of certain *boundary conditions*, these resonators exhibit certain frequency with high order of precision. It is in fact called the resonant frequency. The initial resonant frequency is disturbed when the resonator is influenced by mechanical or electrical interaction, defined as *perturbation*. In other words, it can be said that *'the regime of oscillation of a piezoelectric resonator can be modified by mechanical or electrical perturbations originating from the surrounding medium'*. When the mechanical property like mass changes (because of attraction of chemicals) the boundary conditions gets disturbed and hence modified, resulting in RFS. The RFS can also be achieved by electrical perturbations of boundary conditions. When a dielectric film is coated on the resonator it significantly alters the mechanical boundary conditions. On the other hand, a metallic conducting film alters mechanical as well as electrical boundary conditions.

The devices use piezoelectric material to generate the acoustic wave. Applying an appropriate oscillating electrical power (field) to the piezoelectric material as shown in the figure, can create an oscillating mechanical stress wave because of piezomechanics effect (see Chapter 5). The IDT at input end provides the electric field necessary to displace the substrate. The mechanical wave now propagates through the substrate or on the surface and is then converted back to an electric field, at the IDT on the other side, for read out and measurement of the resonant frequency shift.

Coating enhances the sensitivity of the sensor to specific chemicals. Depending on the application, the resonator can be either coated with the film or can be immersed into the chemical liquid or gas. In either case the loading effect causes damping in the oscillations. That is when a stress-free bulk is brought in contact with the surrounding medium some acoustic energy is drained out from the resonator, resulting in damping of the oscillations. In case of fluid media the sensitivity and hence the performance of the sensor depends on the interface between the quartz resonator and the surrounding medium.

11.4.5 SAW Lines and Design Criteria

The sensitive film is coated along a specific line of the resonator, called SAW line. There exist a large number of SAW lines in a resonator. Since the dimensions of SAW lines are small, numerous lines can be coated with different sensitive films so as to enable the sensor to detect a wide range of analytes. In other words, an array of several coatings can be employed and analyzed to obtain a *signature* of a chemical *fingerprint*. Note that the performance may be affected if the number of coating is not optimized, resulting in cross-link interference.

11.4.6 Integration Advantages

Surface acoustic principle is a very acceptable principle due to the reason that the output, which is simply frequency, can be measured with higher accuracy than any other parameter. Therefore, they are well suited for the design of high-sensitivity sensors.

SAW sensors developed using semiconductors such as Si and GaAs have some potential advantages because the excitation and detection circuitry can be integrated into the same device making the entire sensor smaller and robust. In fact the resonator is integrated with signal conditioning circuitry for instrumentation and signal delivery. The device can be used for real-time detection. The integration system provides excellent sensitivity, ruggedness, direct frequency output and good stability. Such detectors have a wide range of applications including monitoring of the effectiveness of chemical filters,

chemical and biological weapon detection, airport security, and pollution monitoring. Detection of contamination with mass less than 20 picograms/cm^2 is an acceptable typical measurement.

11.15 SINGLE MOLECULE DETECTION

The development of single molecule detection techniques has opened up a new era of life science. The single molecule measurements scheme use combinational techniques including optics, microscopy and electronics to detect single molecules. This is especially useful for protein analysis. For instance, the observation of single molecular interactions of protein molecules can be performed by chemically assembling of a single protein molecule on an electrode. The development involves electronics, microfabrication, bioconjugation and molecular biology. The end effect is the design of molecular sensor.

11.16 SUMMARY

This chapter dealt with the fundamentals of chemical and biological sensors. The ability to monitor and detect various chemical analytes is important to many applications.

The applications can be environmental monitoring, quality control and industrial monitoring. One example is the monitoring of the air quality inside a closed chamber and detection of the presence of toxic or otherwise dangerous gases that may come from spills and leaks. Other important application areas are food processing, perfume, beverage, and other chemical products where complex vapor mixtures need to be analyzed and classified in real-time. The detectors used in these applications are called chemo sensors. Chemo sensors are primarily three types: chemoresistor, chemocapacitor and chemotransistor. These sensors work based on the principle of change in resistance, capacitance and threshold voltage, with respect to the presence of gasses or analytes, respectively. Chemotransistors are very sensitive. The design and construction of chemotransistor is similar to FET (Field Effect Transistor). The other name of chemotransistor is ISFET (Ion-Selective Field Effect Transistor). Mass spectroscopy is used to detect the liquid chemicals effectively. Mass spectroscopy works on the principle of analyzing the *mass spectrum*. The chemical to be detected is first ionized, then passed through a magnetic field to make them available at the detector. Because of the interaction with the magnetic field the ion particles travel along a curved path on a radius that is proportional to the mass to charge ratio (m/z). The curvature will vary for different chemical. A different mass-sensitive device known as fluorescence detector can also detect the chemical species. There exit another spectroscopy method, which is popularly known as calorimetric spectroscopy. The principle is based on the *photothermal signature*. Another popular chemical sensing is achieved by using SAW (Surface Acoustic Wave) sensor, which works based on the principle of *resonant frequency shift*.

Micromechanical and nanomechanical biosensors are devices that measure physical quantities by utilizing variations in the physical properties of specifically fabricated microstructures that originate from biological interactions. Most important biological sensor is the DNA (Deoxyribonucleic Acid) sensor. The detection of DNA is achieved based on hybridization. There are basically three methods by using which DNA can be detected. The methods are (i) optical method, (ii) stress detection method, and (iii) resistance detection method.

Points to Remember

- MEMS devices used for health science and chemical industry are called BioMEMS and chemical microsystems.
- Upon interaction with a chemical species called analytes the physical, electrical or optical properties of the membrane changes.
- Electrochemical sensors change the electrical property such as resistance and potential.
- SnO_2 is used for the detection of hydrogen, carbon monoxide, and nitrogen oxides. ZrO_2 is used to determine oxygen, nitrogen oxide and ammonia.
- Shape and coating sensitive silicon chemical sensing devices are in use to detect combustible gases such as CO, CO_2, NO_x, N_2, and even the H_2O.
- BCMS and MFS (Microfluidic Systems) when integrated in one platform it is called a portable laboratories. The small-sized portable laboratories are known as *chem-labs-on-a-chip* (CLOC).
- Chemoresistor is simply a transducer device in which the resistance of a chemically sensitive layer changes with respect to the amount of chemical absorbed.
- The chemocapacitor is a two terminal device. The principle of operation of chemocapacitor is similar to normal capacitor.
- Ion sensitive chemical sensors are useful in order to detect particular species in surroundings. Basic principle of operation of these sensors also lies with the well-known Field Effect Transistors (FET).
- ISFET is used for the detection of many types of ions however, in each case the electroactive material should be different.
- The ion concentration in the solution influences the gate potential of the ISFET, which in turn modifies the transistor threshold voltage.
- The ISFET based chemical sensors are effectively attractive compared to other conventional types of sensors such as potentiometric or spectrophotometric sensors, as they have good analytical performance, small size and low price.
- Electronic noses (e-nose) are sensors, which mimic the sense of smell.
- Specific fractions of the human DNA reveal valuable data. These data can be utilized for health care and medication purposes.
- The basic element of the DNA analysis system is a DNA probe(s). The DNA analysis system with probe is called DNA sensor.
- The process of recognition is called hybridization. The hybridization reveals the presence of DNA by changing some characteristics of the DNA probe. The hybridization can be achieved in three ways: Optical method, surface stress detection method and electrical detection method.
- In optical method the color of the solution, containing sample DNA, changes.
- Surface stress based hybridization process uses a micro cantilever structure. The cantilever is coated with the detector film that reacts with the biomolecules of interest. A biochemical reaction at the cantilever surface changes in the surface stress. The surface stress causes a bending of the cantilever. A piezoresistor integrated with the cantilever reflects the bending variable.
- The electrical hybridization process uses microelectrodes to which DNA probe is attached. At resting state the electrical resistance between the two opposite electrodes is infinite. Upon hybridization, the resistance falls.
- Fully integrated electronic DNA sensor arrays on CMOS are available in the form of DNA chip.
- Mostly, mass-sensitive sensory system utilizes the principle of *spectroscopy*.
- Electromagnetic (EM) radiation at some frequency is allowed to *interact* with the chemical of interest. Then the properties such as amount diffracted, emitted, scattered and absorbed of that radiation is monitored. The frequency of the EM radiation is determined by the energy levels associated with the property of the target chemical sample.

- The chemical to be detected is first ionized. The ionized chemical is then passed through a magnetic field of pre-defined direction. The ions are accelerated and collected at the detector. The particles travel along a curved path on a radius that is proportional to the mass to charge ratio (m/z). The curvature will vary for different chemical, since the ratio vary from chemical to chemical.
- The mass spectra are measured in terms of relative *abundance*.
- A waveguide fabricated in a flat material such as thin film is known as planer waveguide.
- The CalSpec primarily consists of four building blocks such as, a broadband light source, thermal detector array, a monochromator and the electronics read out circuitry.
- The measurable output in a CalSpec is defined in terms of *photothermal signature*.
- The SAW detection method uses a thin piezoelectric membrane, upon which a selective material is coated, vibrating at certain high frequency. The chemicals to be detected get attracted towards the membrane slowing down its vibration. This is called acoustic attenuation. The change in frequency is referred to as *resonant frequency shift* (RFS).
- Regardless of transduction principle all the sensors are always characterized by sensitivity, stability, applicability, portability and cost.

Exercises

1. Expand the abbreviations BCMS, BCMSAE and BCMSSD.
2. Write down the principle of operation of a simple biochemical sensor. What are the various types of primary sensing principles?
3. Prepare a table to show the principle of operation and the corresponding affected parameters with respect to different types of chemo and bio sensory devices.
4. What do you mean by CLOC?
5. What is a chemoresistor? Write down the expression that signifies the variation of the resistance in the presence of gas. Define the terms used in the expression. How can this expression be modified in order to measure humidity? With suitable schematic diagram explain the principle of operation of a chemoresistor.
6. What materials are used for the chemoresistor? Write down the factors which affect the sensitivity of polymer-based active material used as transducer element in a chemoresistor.
7. What is a chemocapacitor? With suitable schematic diagram explain the principle of operation of a chemocapacitor. The change in the capacitance due to absorption of analyte is related to three different activities. Identify them.
8. Explain the principle of operation of the following transistors.
   (a) FET
   (b) Depletion mode MOSFET
   (c) Enhancement mode MOSFET
9. With suitable schematic diagram explain the principle of operation of a chemotransistor. Write down the expression for the threshold voltage of a typical chemotransistor. Define the terms. What are the two methods used to detect the threshold voltage of a chemotransistor that is being modified by the presence of analytes?
10. Write a short note on e-nose. What is the benefit of having arrayed devices in an e-nose sensor?

11. What do you mean by hybridization in a bio-sensor? What is the expansion of DNA? Comprehensively discuss how a typical DNA sensor works? Describe the three methods of detecting the presence of DNA. What are the important applications of DNA sensors?
12. Write notes on DNA chip.
13. Define the term spectroscopy and mass spectroscopy. Discuss the principle of operation of a mass-sensitive chemosensor. Describe the different parts of a MEMS based mass-sensitive chemosensor.
14. Write notes on the following.
    (a) Fluorescence detection
    (b) CalSpec
    (c) SAW

Chapter 12

CNT and Nanotechnology

Objectives

The objective of this chapter is to study the following.

- The scope of nanotechnology
- Nanotechnology materials
- Structure and properties of fullerenes
- Structure of carbon nanotube (CNT). In particular, armchair, chiral and zigzag nanotube structure
- Single wall carbon nanotubes (SWCNT)
- Multi-wall carbon nanotubes (MWCNT)
- Properties of CNTs
- Applications of CNTs
- Hydrogen storage
- Introduction to molecular machines, molecular pumps and molecular gear

12.1 INTRODUCTION

Nanotechnology[1] is a technology that considers designing products at the nano scale. It owes its name to the prefix nano, a Greek word, which means that the component of a product would be a billionth meter dimension (i.e. 10^{-9} m). The product could be a machine, robot or a computer. Nanotechnology is a relatively new technology that studies properties and behaviour of matter in order to build components and products at the atomic and molecular levels.

We may think of the products being used in our everyday life at the atomic scale. Bear in mind that all the manufactured products are made from atoms. The properties of these products depend on how the atoms are arranged, manipulated and configured. For instance, if we rearrange the atoms in coal we

[1] "…Richard P. Feynman (Born in New York City on the 11th May 1918. He studied at the Massachusetts Institute of Technology (MIT) where he obtained his B.Sc. in 1939 and at Princeton University where he was awarded Ph.D. in 1942) is believed to be godfather of nanotechnology. Eric Drexler, who coined the term within the nanoworld, is the undisputed father of nanotechnology. In 1977, while an undergraduate at MIT, Drexler came up with a mind-boggling idea. He imagined a sea of minuscule robots that could move molecules so quickly and position them so precisely that they could produce almost any substance out of ordinary ingredients in a matter of hours. He pioneered the study of nanotechnology, introducing the term in 1986 to describe Richard Feynman's vision of nanomachines building products with atomic precision…."

should be able to make diamond. Similarly, if we rearrange the atoms of silicon with other composites we can make electronic components such as diodes, transistors, thyristors, triacs, and even semiconductors chips for computers. We never characterize these components as being developed by atomic and molecular manipulation method due to the reason that in practice the design method does not actually handle the atoms and molecules. The design scenario of electronic components, MEMS and MOEMS fall under the category of microtechnology. Conversely, a technology that handles the atoms and molecules to construct a component with dimension of nano scale is called nanotechnology. Some of the worldwide nanotechnology research and development areas are nanoscale phenomena, atomic and mesoscale modeling, carbon nano structures and devices, nano composites, biomaterials and systems, bio nano computational methods, fluidics and nano medicine.

Nanotechnology is much more than what will be presented in this chapter. The reason is that the technological advancements in nanotechnology is seen to be less and till date it has not come up as a fully-fledged technology. It is forecasted that we have to wait for another ten years in order to use nanotechnology conformant products and systems in our everyday life.

This chapter is intended to present fundamental knowledge as far as recent developments in nanotechnology is concerned. The specific definition of nanotechnology is not unclear, but the other frequently used synonym of this technology is *molecular manufacturing*.

12.2 NANOTECHNOLOGY MATERIALS

There exist some materials, which can be used for the design of specific structure and components. For example, semiconductors such as silicon (Si), germanium (Ge), gallium arsenide (GaAs), etc are the materials used for the design of microtechnology-based components and products such as semiconductor chips and even MEMS devices. These materials are selected because of their inherent physical, electrical and chemical properties that are suitable for the design of the desirable components. For instance, one of the properties of the semiconductors is that they are light sensitive, generating photo-voltage or change in resistance upon irradiation by light. This property can be exploited to design photodiodes and phototransistors. Other properties of these materials that they possesses negative coefficient of resistance; meaning that a reduction in resistance comes with increase in temperature. This property can be exploited to generate more number of free electrons within the material block by increasing the temperature, hence it can be used as a temperature sensor. So, what it matters is that we are looking for certain materials that can provide us the desirable properties that we are after.

In the context of nanotechnology the researchers have been looking for certain materials that should inherit the desirable properties so that nanoscale components and structures can be built. In this respect carbon is found to be suitable for the nanotechnology-based components due to its inherent desirable properties. Carbon is a unique atom among other elements because of its ability to exist in a wide variety of structures and forms. Note that pure carbon exists in four different crystalline forms: Diamond, Graphite, Fullerenes and Nanotubes. Carbon atom is the basic building element of these crystalline structures. Out of these, fullerenes and nanotubes are found to be useful in the fabrication of nanoscale components.

12.3 FULLERENES

Fullerenes are all-carbon molecules. The most recognized fullerene is C_{60}, but various other molecules exist, such as C_{70}, C_{76}, C_{84} and C_{102}. The discovery of fullerenes has generated tremendous excitement

and opened up a new field of carbon chemistry. The discovery of C_{60} has stimulated a large activity not only in chemistry but also in nanotechnology. It has opened up the new branch of fullerene-technology that studies the new families of molecules. By the mid nineties about 8000 fullerene compounds were known. Fullerenes are also known as *buckyballs* because their structures are like balls, as shown in the Fig. 12.1(a).

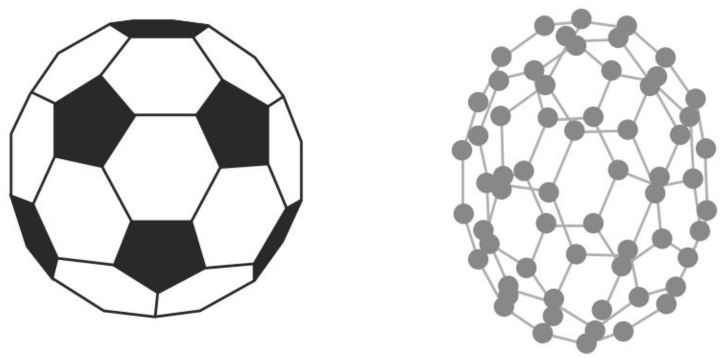


Fig. 12.1 *(a) Shape of the fullerenes C_{60} (b) C_{70}*

C_{60} is a pure molecule that consists of 60 carbon atoms, arranged as 12 pentagons and 20 hexagons. The shape is like a football. This typical spherical shape is based on the physical law, which suggests that a spherical surface is entirely built up from pentagons and hexagons. From the theorem of the mathematician Leonhard Euler it can be stated that a spherical surface can have exactly 12 pentagons and depending on the number of hexagons, molecules of different sizes can be obtained. As a matter of fact fullerenes such as C_{70}, C_{76}, C_{84}, etc. do exist. Scientifically, it can be said that fullerenes are polyeders build up by '*n*' three times coordinated carbon atoms with 12 pentagons and *n* hexagons, where the minimum value for '*n*' is 20. The structure fulfils the Euler's theorem, where a polyeder comprises pentagons and hexagons and has to exactly contain 12 pentagons, to build a closed structure. Fullerenes are apparently large carbon-cage molecules. It is genuine to classify fullerenes as the nanostructured material (NSM) since by itself the fullerenes constitute a physical structure. C_{60} has been assigned the name fullerenes after an American architect Richard Buckminster Fuller who designed domelike buildings that rather resemble the carbon balls. Following features and properties of fullerenes are worth noting.

- The structure of the bucky ball are cage alike and the dimension of the cages are about 7–15 angstroms which is 6–10 times the diameter of a typical atom.
- At much lower temperature of the order of a few hundred °C, fullerene vapor can be formed without breaking the bond. The balls do not break, but they just separate from the solid intact! This property is called sublime behavior. This important property is exploited in growing crystals and thin films.
- The ball can be broken at temperatures of over 1000 °C. So, chemically they are quite stable.
- Fullerenes encompass a very large range of building blocks, which can be the basis for developing new materials with unique properties.
- Adding other functional groups externally or internally can create chemical derivatives of fullerenes. The process of addition is called doping.

- Adding 3 alkali atoms per C_{60} results in a material which goes *superconducting* at a temperature 10–40 ^{o}K.
- Electrical conductivity changes with doping.
- Fullerene based films can function as high temperature lubricants or as super corrosion resistant materials.
- C_{60} possesses strong and weak absorption property in the UV region and over visible region. Thus, they can be used in optical limiting devices for protection of human eyes against laser.
- Fullerenes are insoluble in water and stabile (resistant to change) at air.
- Good solvents for fullerenes are CS_2, o-dichlorobenzene, toluene and xylene. Solubility of C_{60} reaches the maximum near the room temperature.
- The colors of the thin layers of fullerenes vary from yellow to yellow-green.
- The shape of the C_{70} fullerene is rugby-ball-like as it has an extra hexagon (Fig. 12.1(b)).
- Some of the fullerenes shapes are toroidal.

12.3.1 Doping

There is space to intercalate other atoms or small molecules in between the buckyballs since C_{60} has a free electronic band which can hold up to six more electrons. When alkali metal atoms are put into this structure they donate their 1 valence electron to a neighboring C_{60} molecule. The first stable structure is A_1C_{60}. Where, A stands for alkali. It could be Rubidium or Potassium. Addition of alkali is called doping. Further doping enables the production of A_3C_{60}. In this configuration C_{60} molecule now has half full electronic band, so it characterizes a metal and in effect it becomes superconducting at 30 ^{o}K for Rubidium, 20 ^{o}K for Potassium. A_4C_{60} and A_6C_{60} combinations are possible. The former one is a lattice, which is called body-centered-tetragonal fullerenes, and the latter one is called body-centered-cubic fullerenes. A_6C_{60} is the highest Alkali doped fullerene with full electronic band, and it behaves as an insulator.

12.4 CARBON NANOTUBE (CNT)

Extensive research on fullerenes has led to the discovery of other novel molecular structure and constructions. Besides ball-shaped construction, other structures such as longitudinal, X-shaped, Y-shaped and circular (ring) tube are feasible. In this respect a promising group of nanostructured materials is the nanotubes. The centre of nanotube is not necessarily the carbon atoms. Nanotubes are obtained from various materials such as boron nitride and molybdenum. However, when carbons are used as basic building element, the tubes are called carbon nanotubes, CNTs in short. Carbon nanotubes are ultra-fine unique devices, which can offer significant advantages over many existing nanostructured materials (nanotube) due to their remarkable mechanical, electronic and chemical properties. CNTs are apparently superior and has been accepted both from a fundamental point of view and for future applications, as the most important nanostructure because of their unique and interesting properties for the design of a tremendously diverse range of micro- or nanoscale components and product such as electronic components, biomedical devices, nano-composites, gas storage media, scanning probe tips, etc.

Understandably, CNTs is envisaged as graphite sheet rolled into seamless cylindrical shapes as illustrated in Fig. 12.2. The hexagonal lattice of carbon is simply the graphite. When the sheet is rolled, the structure is tube alike and nevertheless it is a single molecule. Each single molecule nanotube is made up of a hexagonal network of covalently bonded carbon atoms. In some cases, the hexagons are

arranged in a spiral form. The sheet appears like a rolled-up chicken wire with a continuous unbroken hexagonal mesh with carbon molecules at the apexes of the hexagons (Fig. 12.2(e)). A CNT has a diameter measuring on the nanometer scale. With strong covalent bonding the CNT has very large aspect ratio (the ratio of length to diameter in this case). CNT are extraordinarily longer in length, usually measuring from about a few tens of nanometers to several micrometers with diameter can be up to 30 nanometers and can be as small as 2.5 nanometers. Apparently, they are hollow cylinders (Fig. 12.2(a–c)); extremely thin with diameter is about 10,000 times smaller than a human hair. Sumio Iijima, a Japanese

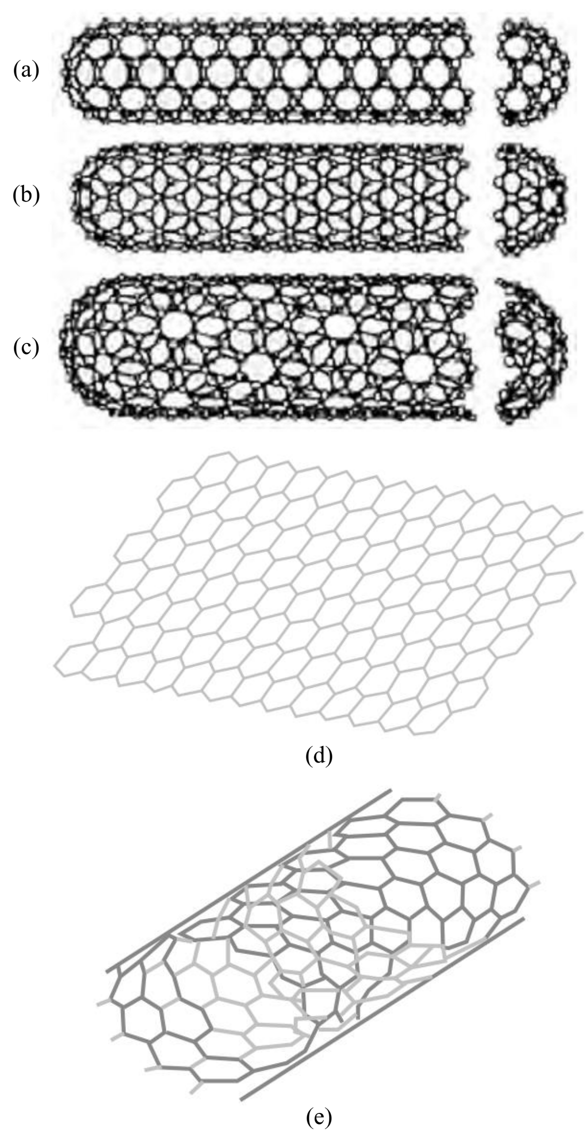


Fig. 12.2 *(a)–(c) Various structures of CNTs, (d) A single layer of graphite called a graphene layer, (e) SWCNTs can be considered to be formed by the rolling of a graphene*

scientist discovered this tiny tubule. Although its structure is tubular but logically, it is a one-dimensional structure.

Carbon nanotubes have many structures, differing in length, thickness, type of spiral, and number of layers. Although they are formed from the same graphite sheet, but their electrical characteristics differ depending on structural variations, acting either as metals or semiconductors. Explicitly, they adhere metallic to semi-conducting properties along with good thermal conductivity. The important properties they possess are;

- Better resilience[2]
- High tensile strength (high strengths and stiffness).
- High current densities (electrical conductivities)
- Good thermal conductivity

Basically, two types of nanotubes can be fabricated. They are:

- Single walled carbon nanotubes (SWCNTs)
- Multi walled carbon nanotubes (MWCNTs)

12.4.1 SWCNT

SWCNT consists of a single graphite sheet with one atomic layer in thickness. MWCNT is formed from two to several tens of sheets arranged concentrically into tube structures as shown in Fig. 12.3(b)&(c). Each of these has its advantages and disadvantages for different applications. Table 12.1 shows a comparison of properties of CNTs with some structural materials.

Table 12.1 *Comparison of properties of CNTs with some structural materials*

Materials	*Young's modulus (GPa)*	*Tensile strength (GPa)*	*Density (g/cm^2)*
Single-walled CNT	1054	~150	
Multi-walled CNT	1200	~150	2.6
Steel	208	0.4	7.8
Epoxy	3.5	0.05	1.25
Wood	16	0.08	0.6

SWCNTs exist in a variety of structures corresponding to the many ways a sheet of graphite can be wrapped into a seamless tube. SWCNT is essentially single layer of pure-carbon atoms rolled into a seamless tube capped at each end by half-spherical fullerene structures. The overall shape remaining cylindrical, rolling of lattice at different angles forms a twist or spiral in the molecular structure of the nanotube. The twist is expressed in terms of what is known as *chirality.* The SWCNT has three basic geometries, namely armchair, zigzag and chiral (shown in the Figure 12.3(d–e)). Each structure has a specific diameter and chirality. Chirality is a measure of wrapping angle. Quantitatively, chirality is described by *chiral angle*. Chiral angle is defined as the angle between the axis of the hexagonal pattern and the axis of the tube. A thirty-degree roll produces an armchair pattern and a zero degree roll forms a zigzag. Any intermediate angle between 0 and 30° makes a chiral nanotube. Armchair and zigzag refer to the pattern of C–C bonds along the circumference of the tube.

[2] The ability of a material to return to its original shape or thickness after being compressed. Glasswool has excellent resilience. (*Source:* www.insulationsolutions.com.au/specifiers/Jargon/)

The chiral angle determines whether the tube is metallic or semiconducting. The armchair structure, with chiral angle 30°, has metallic character. The zigzag tube, where the angle is 0°, can be either semimetallic or semiconducting, depending on the diameter. Nanotube with chiral angles intermediate

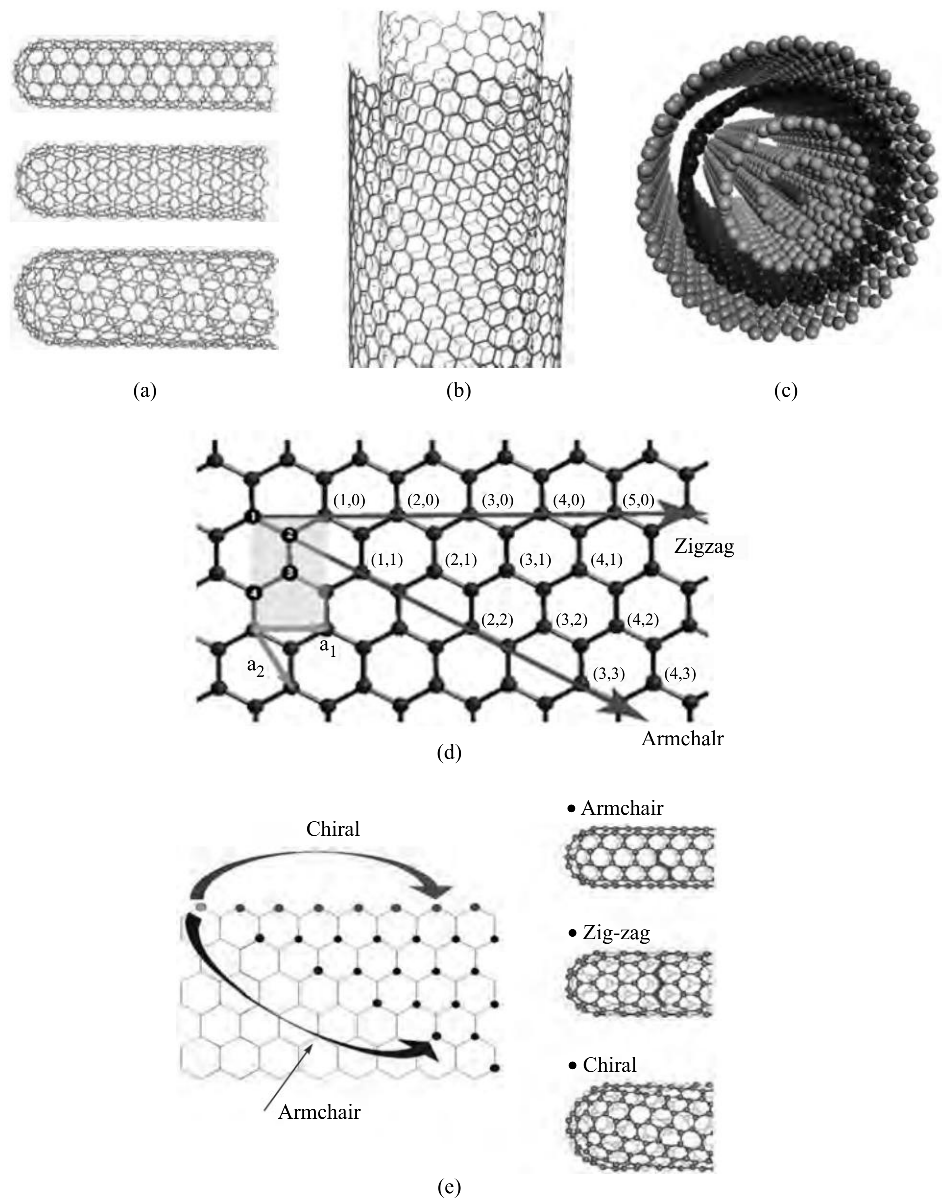


Fig. 12.3 *(a) SWNT—from top is the armchair, zigzag and bottom chiral form; (b) double wall helicoidal form typifying MWCNT (Source: Ebbesen; Poole—JW& Sons); A single strand of carbon atoms (center) is contained in a MWCNT (Source: Zhao et al.); (d–e) Chirality in nanotubes (Source: Kent Seaburn, Nanopedia available at http://nanopedia.cwru.edu/PrintPage.php?page=nanotube.chirality*

between 0 and 30° can behave as both semimetal and semiconductor. SWCNTs differ from MWCNTs in that all of their atoms form a single covalently bound network giving more distinctive electronic and optical properties. The important characteristics and properties of SWCNT are as follows.

- The electronic characteristic of *one-dimensional along-axis* type periodic structured SWCNT is dependent on the way the atoms are dispersed.
- Electronic transport confinement in radial direction is maintained by the monolayer thickness of the nanotubes.
- SWCNTs are metallic or semiconducting, depending on chirality and diameter.
- It shows discrete conductance because of varying band gap
- Maximum current density is approximately 10^{10} A/cm^2. The value is about 1000 times greater than copper wire.
- The nanotube has extremely low electrical resistance.
- In the conductive mode, the conduction is based on what is known as ballistic transport (BT). BT is a current transport mechanism in which the very same electron that enters one side of the tube appears at the other side. Unlike transportation mechanism in semiconductor materials, in CNT however, the electrons in the conduction band of the carbon atom are not considered as the charge carriers. In ballistic transport the resistance of the CNT is independent of its length.
- A single CNT with a natural junction behaves as a rectifying diode
- CNTs are exceptionally strong structures. SWCNTs are stiffer than steel. The tensile strength is more than twenty times greater than steel. They can be bent at sharp angle without damage.
- Maximum strain is approximately 0.11% at 1 V.
- Young's modulus is about 1.8 TPa.
- Gravimetric surface > 1500 m^2/g
- Diameter: 1–100 nm; Length: up to millimetres
- The one-dimensionality of CNT tends to fall out of the composite material
- Thermal conductivity ~6 kW/km.
- CNTs can withstand to temperatures, that is, three times higher than the melting point of copper.

12.4.2 Chiral Vector and Chiral Angle

It has already been mentioned that chirality should be understood as the angle at which the sheet is being rolled. Further, it was pointed out that the ranges of angle vary from 0 to 30 degrees. Figure 12.4 shows an un-rolled lattice that is ready to be rolled. It can be rolled starting from 0 to 30 degrees. In order to define a nanotube or alternatively a nanotube can be specified in terms of matching the points on the left, with the points on the right of the sheet. Once matched, we can connect these points with a vector, which is called *chiral vector* $\boldsymbol{C_h}$. This vector can specify the basic structure of a CNT. In this particular example, as shown in the Fig. 12.4 the vector OB gives the direction of the nanotube axis. The lattice can then be rolled to form a cylinder by matching the points so that the atom at the point O will appear on the top of the point A and the atom at the point B will appear at the top of the point B. In this case OA is the chiral vector. Chiral vector is a part of the nanotube and is perpendicular to the nanotube axis. The vector is determined by the real space lattice translational vectors a_1 and a_2 as shown. This leads to the following equation for the chiral vector $\boldsymbol{C_h}$.

$$\boldsymbol{C_h} = n\boldsymbol{a}_1 + m\boldsymbol{a}_2 \equiv (n, m) \qquad (12.1)$$

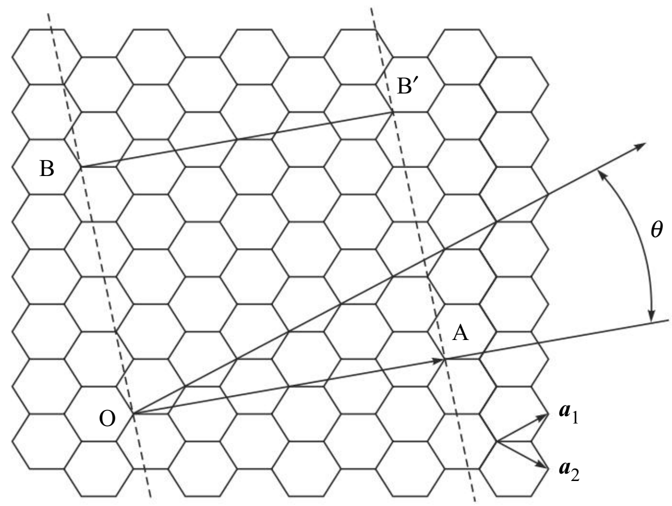


Fig. 12.4 *An illustration showing Chiral vector (Source: Iain Grace, U. of Salford)*

where, n and m are the pair of integers which signifies the position of carbon atom on the two-dimensional hexagonal lattice. Depending on the value of the chiral vector, carbon nanotubes are classified as chiral, zigzag, armchair. The rolling up of the lattice as the placement of the atom at (0,0) on the atom at (m, n) will give nanotubes of different structures as,

$$(m,\ n) = \begin{cases} (m = n) & \text{armchair nanotube structure} \\ (m, n \neq 0 \text{ and } n \neq m) & \text{chiral nanotube structure} \\ (m, n = 0) & \text{zigzag nanotube structure} \end{cases} \tag{12.2}$$

The diameter, D of the nanotube can be determined from the indices *(m, n)*, which is given by,

$$D = \frac{\sqrt{3}a_{c-c}(m^2 + mn + n^2)^{1/2}}{\pi} = \frac{C_h}{\pi} \tag{12.3}$$

where $a_{c—c}$ is the C—C bond length (1.42 Å), and C_h is the length of the chiral vector. The chiral angle is calculated by using the following formula.

$$\theta = \tan^{-1}\left[\frac{\sqrt{3}n}{(2m+n)}\right] \tag{12.4}$$

12.4.3 Structures of CNTs

For various applications, different structural constructions of SWCNT and MWCNT can be developed through appropriate method. For example, continuous sputtering of carbon atoms from the nanotubes lead to dimensional changes. The dimensional change facilitates surface reconstruction with annealing. The reconstructed surfaces eventually appear in various structures. Some typical structures provide typical voltage versus current relationship, which is of great interest. Following structures have been discovered so far.

- Helical
- Bamboo-like structure

- Ring structure
- Cone shape end caps
- X-shaped, Y-shaped and T-shaped CNTs

The helical structure of CNTs is a rope-like structure similar to the double helical DNA structure (Fig. 12.5(a)). Two closed-tip tubes with one tube twisting over the other are observed. Bamboo-like structured CNTs with regular compartment length of approximately 200 nm are seen (Fig. 12.5(b)). These structures are closed tip type with no encapsulated catalytic particles. Note that CNTs are grown by the process in which catalysts are mostly used. Sometimes the catalyst particles are encapsulated in the tube of other structures, but in bamboo-like structures no such encapsulation is observed. CNT rings of typically 0.5 μm diameter are observed. The ring formation can be interpreted as single turn coils with a short overlap between the beginning and end. Figure 12.5(c)–(d) show two different forms of ring-type CNTs. The rings are formed by what is known as buckling effect. A typical cone-shaped CNT is shown in Fig. 12.5(e). Three types of cone-shaped CNT structures exist. The vertex (i.e. the cap of cone) could be either along the axis of the tube, along a line of the circumference or in between the axis and circumferential line. Figure 12.5(e) shows the first type. The vertex of the second type is just like the tip of a pencil. From the standpoint of the chemistry, it is conceptually useful to divide the cone-shaped CNTs into two regions: the end caps and the sidewall. It is essential to know that one structure of the nanotube is seen in other structures. For example, in some bamboo-like nanotubes cone shaped end caps are also seen.

X-like junction with diverse angles between the branches can be formed by CVD (Chemical Vapour Deposition) method. Under careful irradiation one of the branches of the X-junction can be removed, creating Y (Figure 12.5(f)) and T-like junction. Y-shaped CNTs can be elliptical-type or junction-type. Both the types are similar but in the former case the arms are shorter than the latter type. The elliptical-type CNTs are again two types. The differentiating parameter is the length of the arm. The length of the arm could be same as shown in Fig. 12.5(g) or dissimilar as shown in Fig. 12.5(h). The X-, Y- or T-shaped junction tubes can be used for a variety of applications including signal transmission applications. The characteristic curves like current versus voltage of different sets of junctions show robust rectification properties giving rise to the possibility of using these junctions as nanoscale three-point transistors.

12.4.4 MWCNT

As name indicates, a MWCNT has more than one layer of tubes where the tubes are arranged in a coaxial fashion, as shown in Fig. 12.3(b). A MWCNT is considered to be a coaxial assembly of cylinders of SWCNTs one within another. Some MWCNTs show metallic characteristics and other shows semiconducting properties. Various experiments and simulations have estimated that about one-third of the nanotubes are metallic and two-third semiconducting as a result of the tube diameter and the chiral angle between its axis and the zigzag direction (described later). The electrical properties of individual tubes have been shown to vary strongly from tube to tube. The tubes may not be concentric throughout their length. Therefore, the electrical properties are characterized by the amount of disorder and the localization, i.e. the segment at which the electrical properties are being measured. On the other hand, SWCNTs have structural uniformity. Multiwalled nanotubes are being used for the storage of fuel such as hydrogen and methane. It is, therefore, interesting to infer that carbon nanotube is emerging as a building block for nanotechnology, nanoelectronics, nanomanufacturing and nanofabrication. Regardless of their geometrical structure (SW or MW), these tubes are one-dimensional objects with a well-defined direction along the nanotube axis that is analogous to the in-plane directions of graphite.

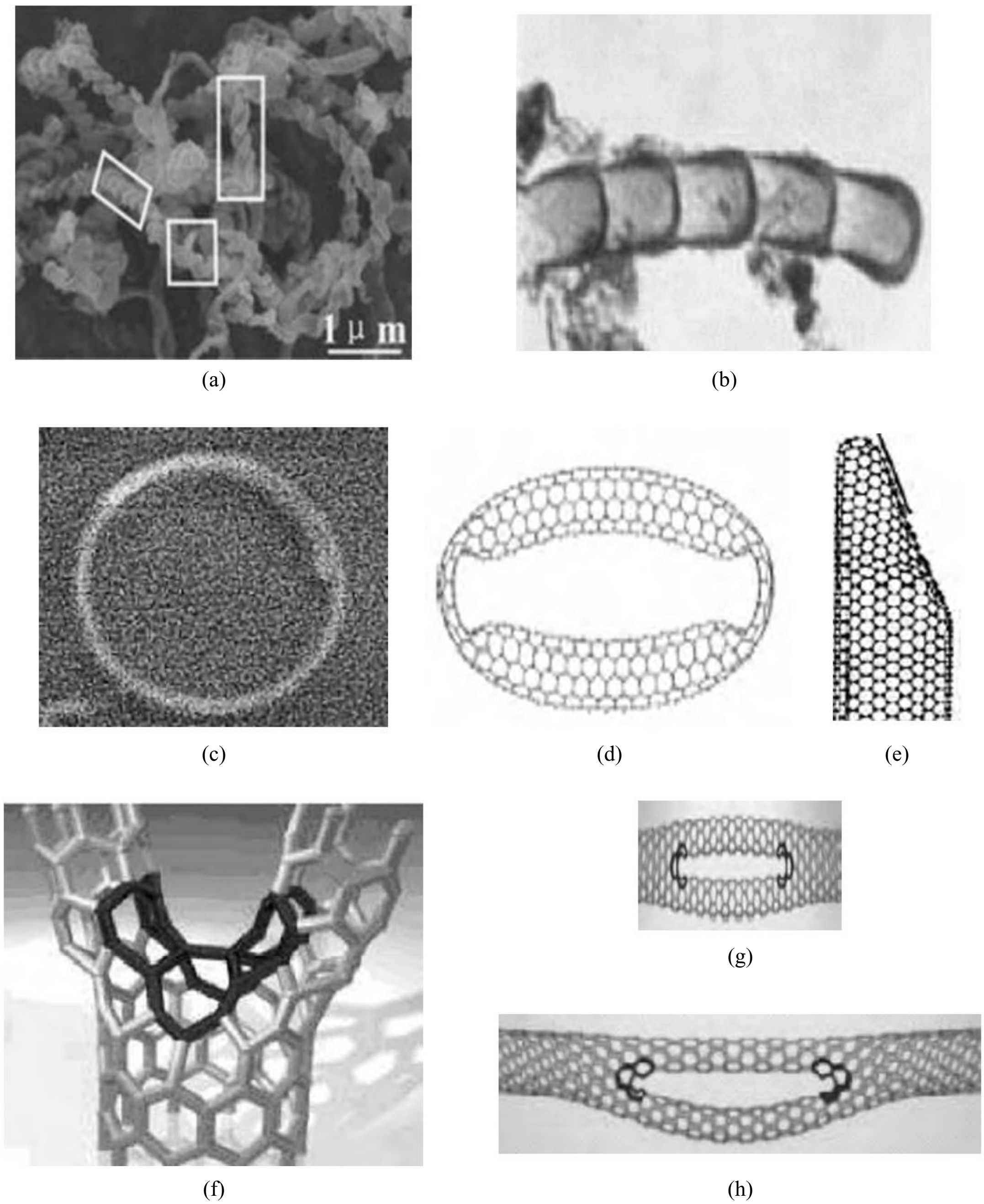


Fig. 12.5 *(a) Helical CNTs, (b) Bamboo-like structure, (c) Ring-type CNT, (d) Another ring-type structure, (e) Cone-shaped (f) Y-shaped junction-type CNT; (g) Y-shaped elliptical-type CNT (arm length same and shorter); (h) Y-shaped elliptical type (different arm length and longer). (Source:* Chen and Huang, Springer, 2005*).*

12.5 DEVELOPMENT OF CNTS

CNTs are developed primarily by three methods.

- Laser ablation
- Electric arc discharge
- Chemical Vapour Deposition (CVD)

12.5.1 Laser Ablation Technique

SWCNTs with uniform narrow diameters ranging from 5–20 nm can be obtained by using laser ablation method. The method (Fig. 12.6) involves the use of a laser beam. Laser beam makes it possible to vaporise a target of a mixture of graphite and catalyst placed inside the furnace. Metals such as Cobalt (Co) or Nickel (Ni) can be used as the catalyst. The temperature of the furnace is maintained at approximately 1200 °C. An inert gas, such as argon, is allowed to pass into the chamber in a controlled manner. The nanotube deposits are recovered at a water-cooled collector, which is kept at a much lower temperature.

The efficiency of the production process is measured in terms of conversion rate of graphite. In this respect laser ablation technique possesses high yield with graphite conversion greater than 85%. The presence of catalyst atoms does not allow the growth of fullerene, instead a selective growth of carbon nanotube is achieved. The reaction temperature can control the tube diameter.

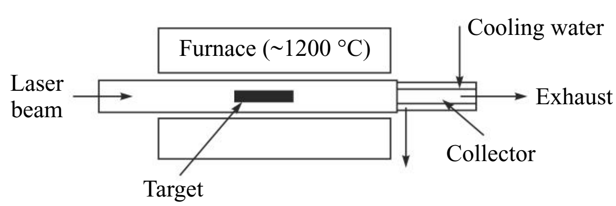


Fig. 12.6 *Laser ablation method of CNT production*

12.5.2 Arc Discharge Method (Fig. 12.7)

The arc discharge method is essentially used to produce nanostructures from high temperature plasma at about 3700 °C. The method is also called DC arc discharge (DCAD) method as the set up uses a pair of electrodes: anode and the cathode, which are kept at a potential difference of approximately 20V DC. The electrodes are kept in a noble gas environment. Helium or argon can be used. Nanostructures with uniform narrow diameters ranging from 2 to 30 nm diameter and 1 µm length can be obtained by using this method.

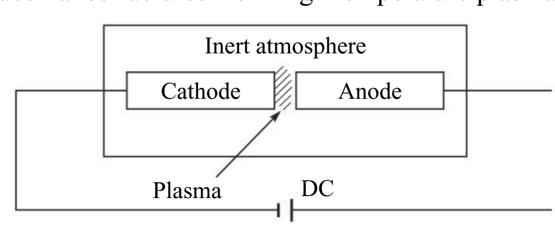


Fig. 12.7 *DC arc discharge method*

DCAD method promises yield with graphite conversion approximately 75%. Nanotubes deposition rate is approximately 1 mm/min. In some cases transition catalyst metals such as Co, Fe or Ni are incorporated into the electrodes in order to favour nanotubes formation against other nanoparticles. When catalysts are used the arc discharge unit is provided with cooling mechanism.

12.5.3 CVD

Chemical vapour deposition (CVD) method is very popular because of its potential for scale up production. In this method CNTs are grown from the decomposition of hydrocarbons at temperature range of 500 to 1200 °C. They are grown on substrates like carbon, quartz and silicon or on floating fine catalyst particles like Co, Fe, Ni from numerous hydrocarbons such as benzene, xylene, natural gas or acetylene. The schematic diagram of a CVD system is shown in Fig. 12.8. It has a horizontal tubular furnace. The furnace is also called reactor. The growth of the nanostructures can occur either in the heating zone, before or after the heating zone. The growth process takes about 30 minutes. The flow of hydrogen gas is typically maintained at 200 ml/min. Argon gas is used to cool the reactor.

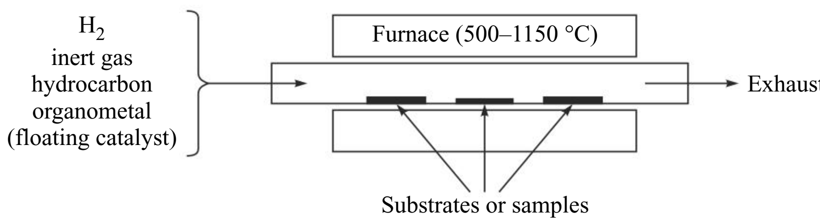


Fig. 12.8 *Catalytic CVD method of CNT deposition*

12.6 APPLICATIONS OF CNTs

CNTs are called supergiant fullerenes and sometimes referred to as carbon nanopolyhedra. After its discovery CNTs have attracted overwhelming number of applications. Further new applications are being predicted due to the extraordinary properties they inherit. Most of the potential applications are, therefore, due to their remarkable properties. For instance, they have internal cavity large enough to encapsulate other materials of the order of 1 to 50 nm in size. This section deals with such potential applications of CNTs. The impending researches that are being carried out all over the world on nanotechnology include study and development of:

- Nanometer-sized semiconductor devices such as electronic devices, FETs, interfacings and probes.
- Energy conversion devices such as supercapacitors
- Lithium ion batteries
- Fuel cells
- Radiation sources such as electron sources
- Hydrogen storage media
- Chemical and biological sensors
- Biomedical application
- Artificial muscles
- Actuators, oscillators
- Nano-machine components such as gears, bearings, mechanisms, etc.
- Nanopumps and pistons

12.6.1 Quantum Wire

Transport characteristics of CNT have shown a tremendous impact in the electronic technology in terms of advocating nanoscale electronic devices. The tube, being characterised by one-dimensional nanoconductor behaves as single electron charging molecular quantum wires. The quantum wire concept is understood as the confinement of electron gas in a 2-dimensional arrangement where the electrons are only free to propagate in one direction, i.e. along the length of the wire. This molecular quantum wire has very long aspect ratio (1:1000) with conduction along the axis of the tube and confinement of electrons around the circumference of the tube. Conductance of an individual carbon nanotubes increases with bias voltage. The V–I curve is staircase in nature. The bias voltage is expressed in mV and the resulting current is in nA. Nobel prize winner Smalley and co-worker have shown for that a bundle of SWCNTs with bridging contacts separated by 200–500 nm, the V–I curve exhibits strong suppression of conductance near zero voltage for temperature less than 10 °K. The linear conductance response of the bundle as a function of the gate voltage showed peaks separated by regions of very low conductance. In an electrical resistance experiment involving a single MWCNT and zero magnetic field, the conductance decreased logarithmically with decrease in temperature. Note that the transport phenomenon of nanotubes displays different patterns depending on the nanotube package. Nanotube packages are considered to be four types: single rope SWCNT, bundle SWCNTs, single MWCNT and bundle MWCNTs.

12.6.2 Transistors

Computers are composed of logic circuits in which the basic building blocks are transistors. A transistor inherits switching behaviour. The extremely thin lightweight nanowire of carbon nanotube, which exhibit metallic-, semiconducting- and insulating characteristics, can become an excellent material for smaller switches and therefore can be integrated together in order to fabricate smarter and lighter computer chips. CNTs superiority over the conventional semiconductor material is reflected through low heat resistance even at diameter of 2 nm, whereby variable currents can be carried at different configurations of interconnections.

CNT transport phenomenon through two platinum (Pt) electrodes (source and drain) on a SiO_2 layer over a p^+ silicon (Si) substrate (Shown in Figure 12.9(a)) shows field effect transistor (FET) characteristics. When voltage is applied to the p^+ gate electrode nanotube is switched from a conducting to non-conducting, acting as an insulator at room temperature. Typical V-I curves of the carbon nanotube FET, where varying the gate voltage from positive to negative, changes from large gap of insulating nonlinear to strongly metallic characteristics. A small voltage applied to the gate changes the conductivity of the nanotube up to a factor greater than 10^6. This magnitude is comparable to the conventional silicon FETs, and the switching time of the device can be estimated to be very fast, almost 10^4 times faster. Figure 12.9(b) shows the V-I characteristics of CNT FET. For higher negative gate voltages the drain-source current saturation occurs indicating that there exists resistance in series with the SWCNT. The gating polarity indicates p-type conduction of the SWCNT.

A schematic diagram of a nanotube FET is shown in Fig. 12.9. The SWCNTs are placed between the source and drain. In some cases, nanotube ropes (nanotube package) are used. Although the combination of their electronic properties and dimensions make carbon nanotubes ideal building blocks for molecular electronics however, it requires assembly strategies for precise localization and interconnections. Table 12.2 describes the performance of CNT FET with silicon based MOSFET.

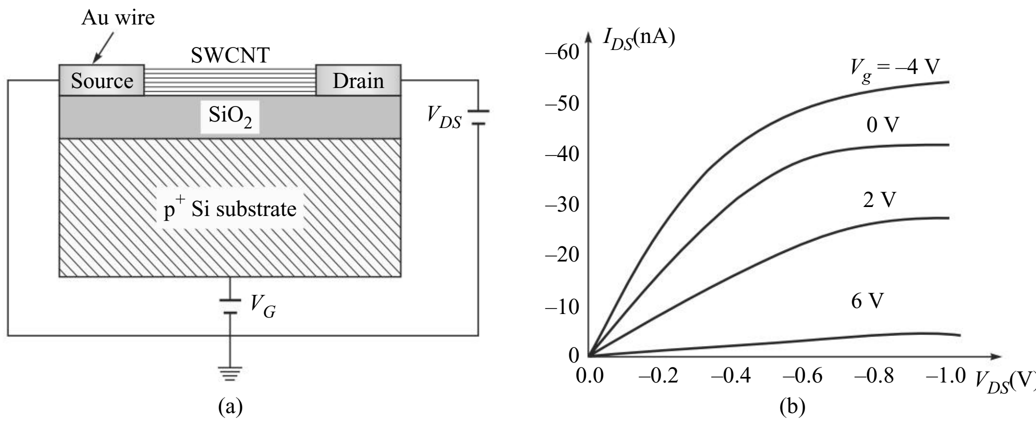


Fig. 12.9 *(a) A schematic of Nanotube FET (Source: Keren, Science, 302, 2003). (b) V-I characteristics of a typical nanotube FET showing p-type behavior (Source- Liu, App, Phy., Let., 79/20, 2001).*

Table 12.2 *Performance comparison of CNTFETs with Si-MOSFETs*

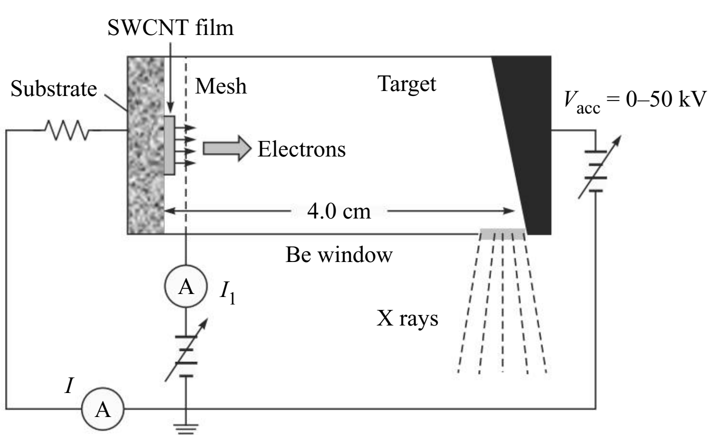
Parameters	*p-CNTFET 1.4×10^{-3}nm (1 V)*	*p-CNTFET 3×10^{-3}nm (1.2 V)*	*MOSFET 1×10^{-4} nm (1.5 V)*	*MOSFET 10 nm (1.2 V)*	*MOSFET 14 nm (0.9 V)*
Drive current Ids (mA/µm)	2.99	3.5	1.04 nFET 0.46 pFET	0.450 nFET 0.360 pFET	0.215 pFET
Transconductance (µS/Am)	6666	6000	1000 nFET 460 pFET	500 nFET 450 pFET	360 pFET
S (mV/dec)	80	70	90	125 101	71
On-resistance (Ω/Am)	360	342	1442 nFET 3260 pFET	2653 nFET 3333 pFET	4186 pFET
Gate-length (nm)	1400	2000	130	10	14
Normalized gate-oxide (1/nm)	80/1 = 80	25/8 = 3.12	4/2 = 2	4/1.7 = 2.35	4/1.2 = 3.33
Mobility ($cm^2/(Vs)$)	1500	3000	-	-	-
I_{off} (nA/µm)	-	1	3	10	100

12.6.3 Biomedical Application

CNTs also possesses properties that can lead to develop biomedical diagnostic devices such as X-ray tube. CNT-based systems can emit sufficient X-ray flux for diagnostics, imaging and photography. X-ray equipment is a device widely used in the medical and industrial applications. The device take tissue photographs of tumor, skeletal fractures and deformations for diagnostics and analysis. They generate high frequency, short-wavelength, high-energy electromagnetic waves that penetrate the body. The conventional thermionic cathode-based X-ray tube has a metal filament that is resistively heated to temperature over 1000 °C to emit electrons, which is in turn targeted and bombarded on a metal anode to emit X-rays. The high temperature requirement in thermionic cathode is actually a limitation. This can be overcome by generating X-ray through field-emission means by the use of CNT materials at the cathode terminal. The field-emission system operates at room temperature.

Figure 12.10 shows the schematic diagram of a typical field emission system. A metal cathode is coated with SWCNTs. The gate electrode is a tungsten mesh that is placed 50–200 μm apart from the cathode. A relatively low voltage between the gate and the cathode is applied. This in effect produces electrons from the cathode. The generated electrons are accelerated and bombarded on a slanted copper target to produce X-ray beam. A high voltage is usually applied between the gate and the anode. Typical value of emitted current that is produced from the metal cathode coated with CNT is about 30 mA. The beam is directed towards the Be window since the target copper is usually kept in slant position. The X-ray flux is sufficient to image humanoid parts. Some of the advantages of the CNT-based X-ray device over the conventional thermionic-based X-ray tube are,

- Prolonged life span of the X-ray tube
- Significant size reduction of the device
- Focus electron beam with programmable pulse width and repetition rate
- Low operational temperature



(a)

(b)

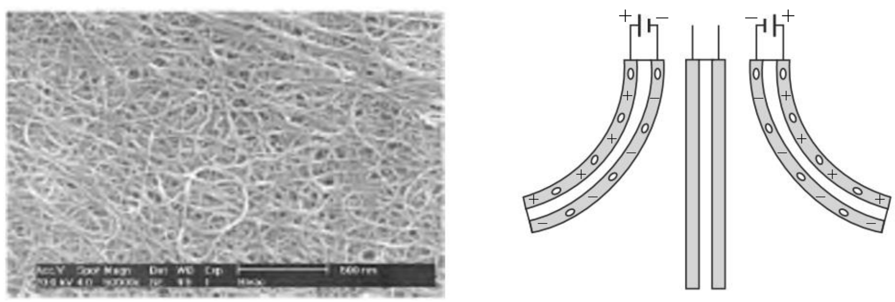
Fig. 12.10 *Field emission-based X-ray device (U. of North Carolina) (a) Schematic of the CNT-based field emission x-ray emitter, (b) X-ray image of humanoid fingers (Source:* Cheng, C. R. *Physique,* Elsevier, 4, 2003*)*

12.6.4 Artificial Muscles

Single walled carbon nanotubes deform when electronically charged. This implies that CNTs possess the characteristics of an actuator that can convert electrical energy to mechanical energy. Baughman and co-worker demonstrated the actuation property of CNTs and characterise them as artificial muscles.

To form a CNT sheet, billions of submicroscopic CNT fibers are tightly packed into bundles with most of their axes remaining in the horizontal plane. Keeping the thickness in the range between 25–50 μm, the bundles can be grown in horizontal planes to make a sheet-like structure. The sheets are simply the bundles of SWCNTs and are referred to as *bucky paper* (Fig. 12.11(a)). Such sheets are the basic building block of the artificial muscle. When two sheets are connected to a battery they perform mechanical work similar to natural muscle. To observe the muscular actuation property, the bucky

papers are to be placed in the sodium chloride electrolyte solution. Upon applying few volts, significant deflection in one direction can be observed. Changing the polarity of the voltage, one can reverse the deflection direction (Fig. 12.11(b)). Further, if an AC input is applied, the bucky paper can oscillate. This oscillatory behavior can also be exploited for many other actuating applications. The benefit of CNT-based artificial muscle is that it has ability to generate large forces utilizing simple electrostatic actuation technique as well as has the ability to operate at low voltages and even at extremely high temperatures.



(a) (b)

Fig. 12.11 *Actuation properties of bucky paper (a) Image showing the nanostructure (Bucky paper), (b) actuation characteristics of the artificial muscles displaying positive or negative deflection when voltage was applied (Source:* Baughman, Science, 1999*)*

12.6.5 Fuel Cells

In general a fuel cell is an electrochemical system that generates electricity by converting chemical energy into electrical energy. Justifiably a fuel cell consists of an anode and a cathode separated by an electrolyte that acts as the ion carrier. An oxidant is fed to the cathode to supply oxygen and a fuel is fed to the anode to supply hydrogen. When hydrogen is used as fuel, it is electro-oxidized. This in effect produces electrons and protons in the anode. The process can be represented as follows.

$$2H_2 \rightarrow 4H^+ + 4e^- \qquad (12.5)$$

Electrons are transported through an external circuit through cathode, as shown in the Fig. 12.12. They arrive at the cathode, protons go through the electrolyte and finally both the electrons and protons are recombined with O_2 in the cathode to give H_2O as expressed below.

$$O_2 + 4e^- + 4H^+ \rightarrow 2H_2O \qquad (12.6)$$

The principle described above is a kind of proton exchange through a membrane coated on the electrode. The exchange phenomenon eventually constitutes fuel cell for which such a system is known as Proton-Exchange Membrane fuel cell or PEM fuel cell. The heart of the PEM fuel cell is Membrane Electrode Assembly (MEA). Refer Fig. 12.13 for detail description on PEM and MEA. PEM fuel cell is one of the currently known five types of fuel cell (alkaline, phosphoric acid, molten carbonate and solid oxide fuel cells) that convert chemical energy of fuels (e.g. H_2 and O_2) directly into electricity, heat and water. An MEA usually consists of a sheet of proton conducting polymer electrolyte membrane with two platinum (Pt) electrodes.

Several challenges to widespread implementation of fuel cell technology are needed, although novel inexpensive and long-lasting electrocatalyst materials are major factors. Perfluorinated sulfonated cation exchanger membrane (Nafion^{TM}) has been commonly used in MEA for PEM fuel cell to provide

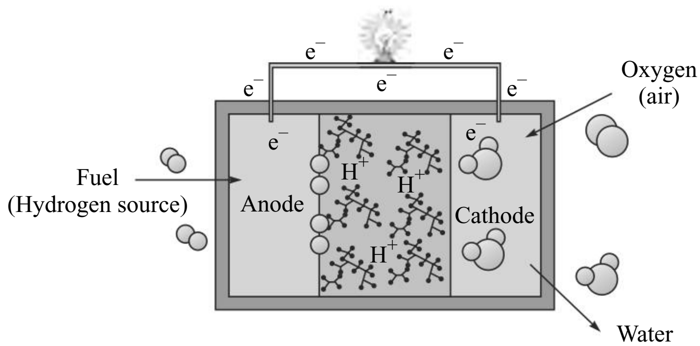


Fig. 12.12 *Proton-Exchange Membrane (PEM) fuel cell (Source:* Mainardi, Springer, 2005*)*

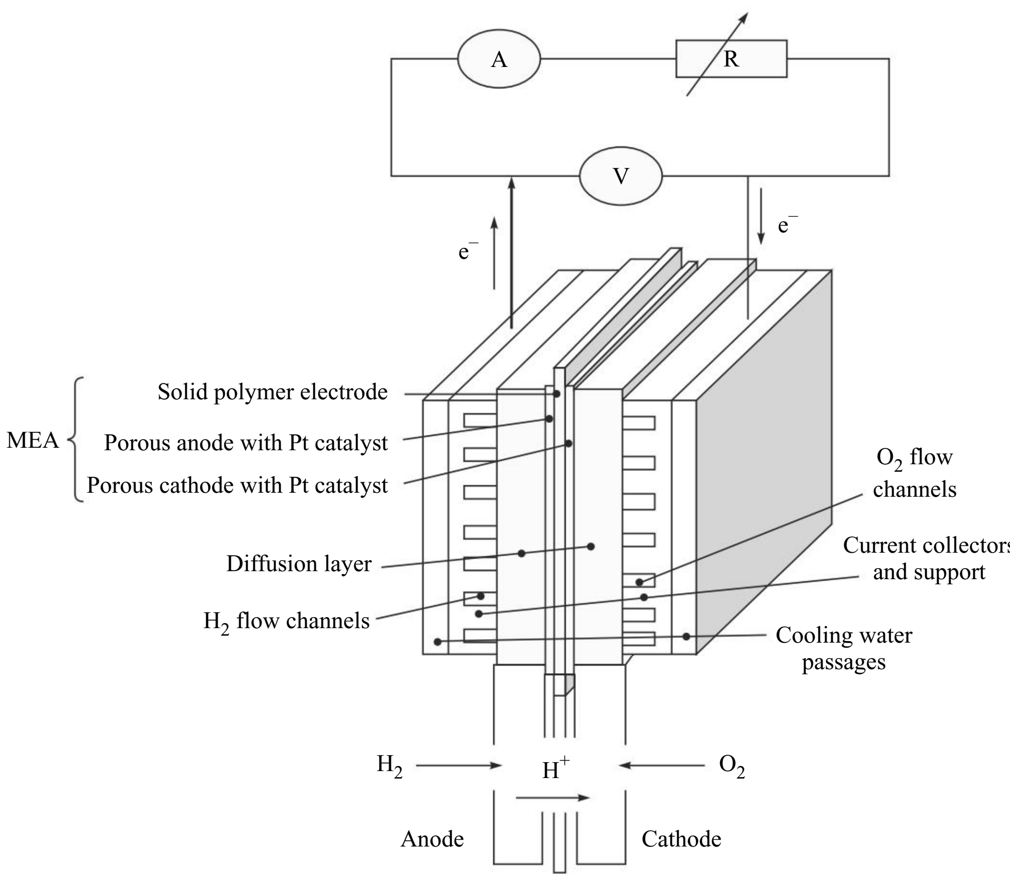


Fig. 12.13 *Schematic diagram of MEA and single cell testing apparatus (Source:* Iyuke, J. *of Power Sources,* 114/2, 2003*)*

mechanical support, insulation and as solid electrolyte for H^+ transport. These qualities can be improved further by the use of appropriate catalytic materials. Current fuel cell cathode catalysts are made of platinum and platinum-based alloys, which exhibit reasonably good catalytic activity for the oxygen

electroreduction reaction (ORR). Also, the type of fuel used dictates the appropriate type of anode catalyst needed. The appropriate electrocatalysis should enhance the reaction rate for oxygen adsorption and breaking the O–O bond during the reduction reaction. For instance, although platinum-based catalysts enhance the ORR, still the oxygen kinetics is significantly slow. The exchange current density of oxygen electroreduction on platinum is about 1.0 nA/cm^2 in acidic electrolyte, which is six orders of magnitude lower than that of hydrogen oxidation. This sluggish oxygen reduction reaction on the Pt catalyst causes a large overpotential at the cathode, requiring a much larger amount of Pt in the cathode than in the anode to achieve the kinetics requirements. Hence, there is a need to enhance the oxygen electroreduction kinetics for which new or improved cathode catalysts must be incorporated. What is required is that the MEA should have higher surface electroactivity in terms of providing improved qualities, better workability including chemical, pressure and heat stability of the membrane coated on the electrodes of the PEM fuel cell. Pt/graphite electrode has been the usual electrocatalytic electrode in most of the PEM fuel cell. However, in the electrochemical and electrocatalytic studies in a PEM fuel cell of a finely and highly dispersed Pt on carbon nanotubes (Pt/CNT) electrode as compared to Pt/graphite electrode shows a remarkable superiority over the latter. The design of new metal catalysts and composites for improving the efficiency of electrode reactions are underway. Carbon nanostructures such as fullerenes and CNTs can be utilized to support to boost the electrode performance of the MEA. Electrophoretically deposited fullerene films are capable of serving as good nanostructured carbon supports. Further, electrophoretic deposition of nanotubes on various electrode surfaces is also a more effective way of depositing carbon nanostructures than the other counterpart. Recent advances in the application of nanostructured carbon-based materials have suggested the possibility of using carbon nanotubes as novel electrocatalyst supports. Studies have shown that Pt nanoparticles supported on carbon nanotubes display remarkably higher electrocatalytic activity.

PEM fuel cells are outstandingly good power source because they are relatively lightweight, can be operated at low operating temperature to provide high power density, and can start-up quickly with rapid response to varying loads. PEM fuel cells are considered as alternatives to batteries in portable electronic equipments, such as laptop computers, cellular phones and other hand-held devices. In the forgeable future fuel cells can be used in our everyday life and can hopefully replace the traditional messy gas and electricity connections providing added advantages of environmental pollution problem such as global warming caused by the exhausts releases from the internal combustion engines.

12.6.6 Hydrogen Storage in CNTs

Scientists are in search of novel material to store hydrogen as fuel for a clean, renewable and environmental friendly technology. In a generalised sense it can be stated that the hydrogen storage for PEM fuel cell can alleviate the petroleum reserves. It has been forecasted that a compact passenger vehicle could be powered by fuel cell that would require 4 kg H_2 for a 400 km driving range. Quite a number of materials have been proposed, but amongst them carbon materials such as carbon nanotube (CNT) and graphite nanofibre are receiving the most attention due to their large capacity to adsorb hydrogen.

12.7 REMARKS ON PROPERTIES OF CNTs

Carbon nanostructures have attracted considerable attention due to their special molecular, electronic, optical and mechanical properties.

- CNTs are quasi-one-dimensional cylindrical aromatic macromolecules and have shown to be chemically inert.

- The curvature-induced pyramidalization and misalignment of the π-orbitals of the carbon atoms induces a local strain.
- Carbon nanotubes are expected to be more reactive than a flat graphene sheet.
- The end caps are always quite reactive, irrespective of the diameter of the CNTs.
- CNT becomes either metallic or semiconducting depending on its chiral vector, i.e. boundary conditions in the circumference direction.
- An electron in a nanotube is a massless neutrino on a cylinder surface with a fictitious flux called Aharonov-Bohm flux[3].
- The optical activity of chiral nanotubes disappears if the nanotubes become larger.
- CNTs are as stiff as diamond.
- Weak region of mechanical strength of CNT is the center of the tube. The connection region is stronger than the tube itself.

12.8 MOLECULAR MACHINE COMPONENTS

We are well acquainted with macro level machines, which are composed of rotating shafts and gears to achieve either translational or rotational motion. An extremely small gear mechanism using a nanocomposite material of nylon and vapor-grown carbon nanofiber has been developed jointly by Prof. Morinobu Endo of Shinshu University and Showa Denko of Kitagawa Industries Co., Ltd. and Seiko Instruments Inc. in Chiba. The vapor-grown carbon nanofiber used in the composite is now commercially produced with trade name of VGCF. The gear mechanism is the smallest in the world based on the current production technology, measuring 0.2 mm in gear diameter, 0.025 mm in module, 0.09 mm in shaft diameter and having six cogs. The gear mechanism can drive a gear integrated with the second hand of the watch. The nanocomposite material is characterized by the ease of injection molding of microparts due to the mixing of carbon nanofiber, ensuring reproduction of delicate shapes. Although the gear uses carbon nanofiber but this gear mechanism is not a nanotechnology conformant product (not even microtechnology conformant as the dimension is in mm scale) but the example has been given to draw attention of the reader.

Molecular level machines are predicted to do similar job of microscale machineries. The molecular machines are composed of parts and components, whose dimensions are clearly in the nano scale range. Nanotechnology conformant passive nanocomponents such as gear, bearing, rack-and-pinion and so on have not yet been designed, however, many simulation based feasibility studies are underway in many research centers around the world.

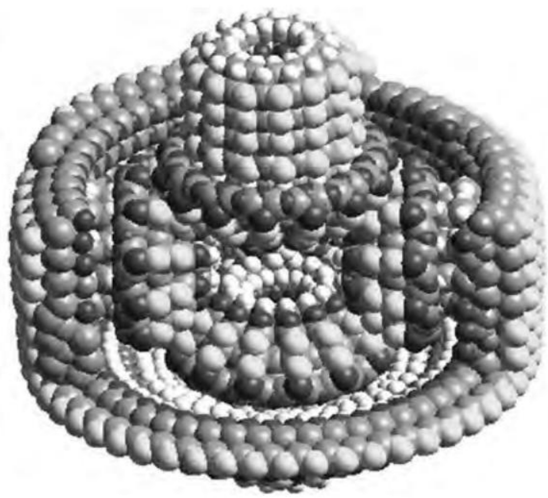
K. Eric Drexler of Institute for Molecular Manufacturing has simulated the design aspects of a typical molecular gear. A simulated differential gear mechanism for a molecular pump was studied by him and is shown in the Fig. 12.14(a). A differential gear links two shafts through a casing. The figure shows that the two cylindrical shafts and their facing bevel (conical) gears, along with two of the four casing-mounted side-gears that mesh with both shaft-gears.

Figure 12.14(b) shows a molecular pump. It is expected that this molecular machine can pump neon (Ne). The left side image shows the chamber wall of the pump containing the pump housing. The right

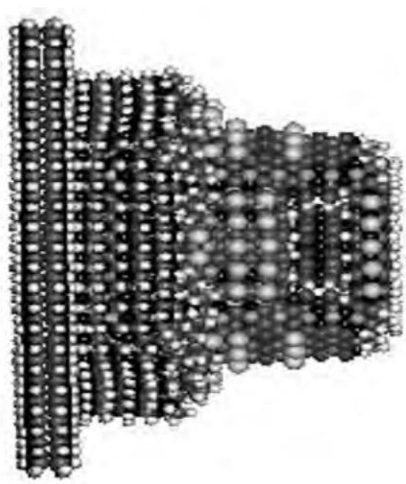
[3] The Aharonov-Bohm effect is a quantum mechanical phenomenon. Aharonov and Bohm predict that a charged particle is affected by electromagnetic fields in regions from which the particle is excluded. The effect demonstrates that the electromagnetic potentials are the fundamental quantities in quantum mechanics. A tube becomes a metal or a semiconductor, depending on whether the amount of the flux vanishes or not.

image is the pump rotor. The rotor has grooved cylindrical bearing surface at both the ends. The middle part of the rotor is a screw-threaded cylindrical segment. While in operation, the rotation of the shaft moves a helical groove past longitudinal grooves inside the pump housing. The pump and segment of chamber wall pictured here contain 6165 atoms.

In another computational study, Han *et al.* at NASA Ames Research Center have suggested that nanotube based gears (Fig. 12.15) can be made and operated and the gears can work well if temperature is lower than 600–1000 °K and rotational energy is less than the teeth tilting energy at 20°.



(a)



(c)

(b)

Fig. 12.14 *(a) A molecular differential gear, (b) Molecular pump, developed by K. Eric Drexler (c) Pump rotor (Source:* www.imm.org, *Courtesy:* Foresight Institute, Institute for Molecular Manufacturing, Xerox*)*

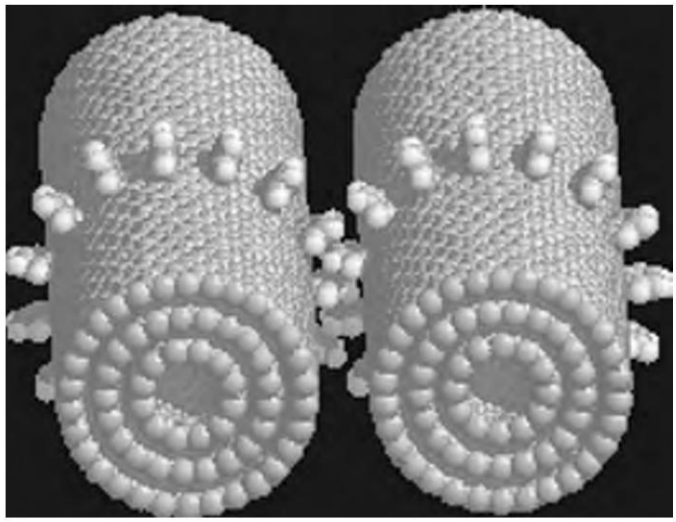


Fig. 12.15 *A carbon nanotube based gear (Source:* Han, 2004*)*

12.9 SUMMARY

Micromanufacturing and nanotechnology possess the same basic goal of promise to change the present and future of manufacturing from the conventional top down to a new bottom up, atom-by-atom technique to achieve precise and specific objects and processes. Carbon nanotube or cylinder of carbon atoms is one of such materials produced atom by atom whose mechanism is yet to be clearly understood. Carbon nanotubes have diameters ranging from less than 1 to 100 nm, and exhibit unique mechanical, electronic and magnetic properties, which have made their research the most active and interesting field in carbon related science in the recent years. This chapter discusses carbon nanotube technology, production and applications since there is presently tremendous interest in their commercialization for both near and future applications.

Points to Remember

- Nanotechnology is a technology that considers designing products at the nano scale.
- MEMS and MOEMS fall under the category of microtechnology. Conversely, a technology that can handle the atoms and molecules to construct a component with dimension of nano scale is called nanotechnology.
- Carbon is found to be suitable for the nanotechnology-based components due to its inherent properties.
- Pure carbon exists in four different crystalline forms: Diamond, Graphite, Fullerenes and Nanotubes.
- Fullerenes are all-carbon molecules. The most recognized fullerene is C_{60}, but various other molecules exist, such as C_{70}, C_{76}, C_{84} and C_{102}.
- Fullerenes are also known as *buckyballs.*
- Fullerenes are polyeders build up by 'n' three times coordinated carbon atoms with 12 pentagons and n hexagons, where the minimum value for 'n' is 20.
- Adding 3 alkali atoms per C_{60} results in a material which goes *superconducting* at quite a high temperature such as 10–40 ^{o}K.

- Nanotubes are obtained from various materials such as boron nitride, molybdenum or carbon. However, when carbons are used as basic building element, the tubes are called carbon nanotubes, CNTs in short. A CNT has a diameter measuring on the nanometer scale.
- Understandably, CNTs is envisaged as graphite sheet rolled into seamless cylindrical shapes.
- The overall shape remaining cylindrical, rolling of lattice at different angles forms a visible twist or spiral in the molecular structure of the CNT. The twist because of the way it was rolled, is expressed in terms of what is known as *chirality.*
- Each structure has a specific diameter and chirality. Chirality is a measure of wrapping angle.
- With strong covalent bonding and with such a small diameter CNT has very large aspect ratio (the ratio of length to diameter).
- Sumio Iijima, a Japanese scientist discovered CNT.
- The two types of CNTs are (i) Single walled carbon nanotubes (SWCNTs) and (ii) Multi walled carbon nanotubes (MWCNTs).
- Quantitatively, chirality is described by *chiral angle*. Chiral angle is defined as the angle between the axis of the hexagonal pattern and the axis of the tube.
- The conduction in CNT is based on what is known as ballistic transport (BT).
- CNT behaves as single electron charging molecular quantum wires.
- An MWCNT has more than one layer of tubes where the tubes are arranged in a coaxial fashion.
- Multiwalled nanotubes are being used for the storage of fuel such as hydrogen and methane.
- The different structures of CNTs are (i) Helical, (ii) Bamboo-like structure, (iii) Ring structure, (iv) Cone shape end caps, (v) X-shaped, Y-shaped and T-shaped CNTs.
- CNTs are grown by the process in which catalysts are mostly used.
- CNTs are developed primarily by three methods, (i) Laser ablation, (ii) Electric arc discharge, and (iii) Chemical Vapour Deposition (CVD).
- Nanotube packages are considered to be four types: single rope SWCNT, bundle SWCNTs, single MWCNT and bundle MWCNTs.
- The current versus voltage curves of different sets of CNT junctions show robust rectification properties giving rise to the possibility of using these junctions as nanoscale three-point transistor.
- CNTs also possesses properties that can lead to develop biomedical diagnostic devices such as X-ray tube. CNT-based system can emit sufficient X-ray flux for diagnostics, imaging and photography.
- Single walled carbon nanotubes deform when electronically charged. This implies that CNTs possess the characteristics of an actuator that can convert electrical energy to mechanical energy, or vice versa.
- CNT-based artificial muscles are referred to as *bucky paper.*
- If an AC input is applied, the bucky paper can oscillate.
- In general a fuel cell is an electrochemical system that generates electricity by converting chemical energy into electrical energy.
- The heart of the PEM fuel cell is Membrane Electrode Assembly (MEA).
- PEM fuel cells are outstandingly good power source because they are relatively lightweight, can be operated at low operating temperature to provide high power density, and can start-up quickly with rapid response to varying loads.
- It has been forecasted that a compact passenger vehicle could be powered by fuel cell that would require 4 kg H_2 for a 400 km driving range.
- Molecular level machines are predicted to do similar job of microscale machineries.
- Nanotechnology conformant passive nanocomponents such as gear, bearing, rack-and-pinion and so on have not yet been designed, however, many simulation based feasibility studies are underway in many research centers around the world.

Exercises

1. What is the approximate dimension of a typical molecule? Nanotechnology is also known as molecular manufacturing. Why is it so?
2. Mention some materials which can be used for the fabrication of nano devices.
3. What do you mean by a buckyball? What does Leonhard Euler theory state? Mention the molecular symbols of some of the fullerenes.
4. What are the noteworthy features and properties of fullerenes?
5. What happens if alkali metals are doped into the C_{60} fullerene?
6. Who discovered carbon nanotubes (CNT)? Give an example how understandably a CNT could be formed? Mention different types of CNTs in the way the graphite sheets are rolled. What are the important properties of CNTs? Distinguish between the SWCNTs and MWCNTs.
7. List out important properties of SWCNTs.
8. Write short notes on the following.
   (a) Chiral vector (b) Chiral angle
9. With schematic diagram show different structures of CNTs and explain their features, respectively.
10. Discuss various methods of fabricating or developing the CNTs.
11. Enlist some of the broad futuristic applications of CNTs.
12. What do you mean by quantum wire? Mention the important discovery of Smalley's work.
13. How can transistors be formed by using CNT materials? Draw the V/I characteristics of nanotube material-based FET (Field Effect Transistor).
14. How are CNT materials used for biomedical applications, especially for X-ray devices. What are the disadvantages of conventional thermionic cathode-based X-ray tubes? How can CNT based systems overcome the problems?
15. Describe the principle of operation of a CNT based artificial muscle? What would happen if DC and AC voltages are separately applied to the bucky papers?
16. Discuss the concept of fuel cell for nanotechnology applications. What is the role of catalysts in the fuel cell? What do you mean by PEM fuel cell?
17. Why is hydrogen storage important? State whether CNT can store hydrogen? Mention some of the applications of stored hydrogen.
18. Remark on some of the properties of CNT.
19. Write notes on the following.
   (a) Molecular machine
   (b) Molecular gears and pumps

Chapter 13

Simulation Based Micro and Nanosystem Design

Objectives

The objective of this chapter is to study the following.

- The simulation tool in general
- The need of simulation tool in MEMS design
- Introduction to FEM
- The design flow within the simulation environment
- Some available simulation tools
- The meaning of multiscale design
- Difference between atomistic and continuum model
- Application of multiscale design approach with regard to MEMS and NEMS
- Concluding remarks

13.1 INTRODUCTION

This chapter introduces simulation based design methodology and the concept of multiscale design approach. Simulation is a design methodology using which designers study the performance of the system (an MEMS structure or component, for example) prior to its real design. Based on fundamental governing laws the simulation study is carried out with the help of a computer.

Multiscale design approach is a unified advanced simulation methodology that treats and simulates each and every element of the MEMS device at different scales lengths depending upon their structure. A wide range of phenomena such as fracture, deformation and microstructural evolution involve processes across multiple length and time scales. Multiscale simulation methodology aims to provide a better macroscopic description of those phenomena by linking simulations across scales in order to incorporate the microscopic-level physics and mechanisms.

13.1.1 Simulation

The development of new devices and their applications depends on having adequate CAD (Computer Aided/Assisted Design) tools for simulation and evaluation. Simulation and evaluation has been used to

explore and test the feasibility of new engineering designs in an optimized way. Utilizing simulation methods for engineering systems has many advantages and there is great deal of validation and optimization. Simulation technologies provide opportunity to the developer and system integrator to model, explore and try out a variety of design strategies in order to advocate matured, viable and effective products. Simulation can improve productivity since the activity brings systems into operation more rapidly. Eventually, simulation plays a major role in the design of MEMS devices.

13.1.2 Multiscale Design Approach

In multiscale design, the objective is to predict the performance and behavior of structured materials across all relevant length and time scales utilizing fundamental physical principles. At the atomic (nanometer-scale) level, electrons govern the interactions among atoms in a solid. They also govern the electromagnetic fields. In order to characterize the collective behavior of the atoms, it is therefore necessary to accept the quantum mechanical descriptions in a material. On the other hand, at the macro-scale (centimeters-scale), forces arising from macroscopic stresses and/or temperature gradients are considered as the controlling elements of materials performance. At the intermittent scales (micrometer-scale), defects such as *dislocations* characterize mechanical behavior. The net outcome of these scale-related interactions is often described in the form of a *constitutive law* that ultimately governs behavior.

The design of MEMS and NEMS (NanoElectroMechanicalSystems) relies on meticulous understanding of the mechanics of the microelement. Atom-by-atom understanding of matter is reflected through atomistic model theory. On the other hand continuum theories describe a system in terms of some physical variables such as mass, for instance. Atom-by-atom understanding and continuum understandings of materials are considered two different scenarios. When dealing with microelements of MEMS and NEMS, neither one (from the above two modeling theories) can be used for analysis. Note that the simulation is carried out based on some fundamental model equations. The model equations used in these two domains (Atomistic and Continuum) are very different and neither one can be used to describe the other since problem encountered here is concerned with what is known as *length scales*. Atomistic and continuum models are themselves under their own scale domains. When the model equations are used in another domain, the assumptions in constitutive law disappear violating the constitutive relations. The atomistic models and theories may not be able to simulate the continuum complexes due to the fact that simulation of large number of atoms is a time consuming task, no matter how many computer resources one throws into the problem. One way to treat the microsystems is to introduce simplifying approximations to both the theories in such a way as to ensure that the resulting model equations are simpler, common and faster to solve and analyse. It is thus essential to develop scale dependent theories, which should bridge the gap between the atomistic and continuum model in order to determine and predict the material properties and mechanical response, respectively, at all scales such as meso, micro, and even nano scales. Thus comes the idea of multiscale design philosophy.

Multiscale design technique is a computational simulation method of the mechanics of materials at the micro and nano scales. The Multiscale Modeling of Materials (MMM) approach has been shown to rely on systematic reduction of the degrees of freedom at natural length scales. The coupling of atomistic models with continuum models is somewhat difficult, since the characters of the models differ. The development of global design methodology employing materials modeling that spans length and time scales from individual atoms to engineering structures (MEMS and NENS structure in fact) emphasises compatibility and interoperability between the models at adjacent scales. The multiscale technique thus

exploits computational materials science and engineering in order to develop a design methodology that incorporates models of structure-process-property relationships of materials into engineering design software, enabling materials design to be an integral part of the global design process.

This chapter has two parts. At the beginning it discusses the need of simulation tool within the MEMS environment followed by description of some commonly used MEMS simulation software such as HFSS, DS and CS MEMS, FEMPRO, ANSYS and SUGAR. In the second part, the chapter introduces the concept of multiscale design methodology and approaches.

13.2 THE NEED OF SIMULATION TOOL

In computing environment, simulation software, as they are referred to, are application programs run as graphic-like platform for the interactive design, programming and evaluation of systems. Simulation, sometimes called virtual platform, provides an important foundation for system development. The platform can be used to carry out design for manufacturability studies including conceptual design, visualization, and evaluation without tying up the real design of physical device or processes. Using accurate mathematical models of the system, design engineers simulate the operation of the system and generate or verify the desired actual physical and task parameters before its real design starts. The virtual design environment can greatly improve reliability and availability. Virtual methods can also help to shorten the design-to-manufacturing cycle by enabling the users to correct errors and to identify optimized design parameters and requirements before they reach the real implementation stage. This phase is known as prototype testing.

The design of any product can benefit from computer-assisted simulation and optimization tools prior to its prototype testing followed by real design. Figure 13.1 shows distinction between the traditional and advanced ways of designing the products. For more than two decades the development of Computer Aided/Assisted Design (CAD) tools has gone hand-in-hand. CAD tools are well accepted in the development of IC processes. SPICE is an obvious example. Simulation tool plays a similar role in the design of complicated MEMS. Simulation tool can handle every aspect of design and development

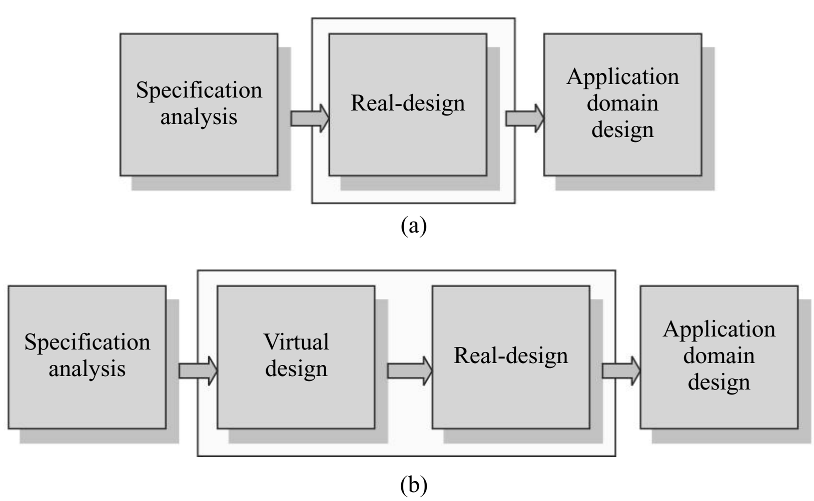


Fig. 13.1 *Traditional versus simulation-oriented design*

process of MEMS devices and systems. The goal of the simulation is to close the design loop by enabling design, fabrication, comparison of measurement with other data, diagnostics, and then re-design. The simulation can improve the capability by reducing the time of the design process.

Simulation is more than just a technology as it handles the optimality globally. It studies the system behavior while considering the system as more than the sum of the individual components. Through simulation software, the system can be segregated for individual study and can also be aggregated for studying the combined effect. Simulation thus can provide insights into the designs of, for example, processes, architectures, properties, effects, etc. before significant time and cost have been invested. Simulation is being increasingly emphasized in the MEMS domain, where there is evidence that its impact on costs, quality and reliability is non-trivial. Some of the simulation software used for MEMS applications is described below. Prior to that it may be necessary to introduce the concept of Finite Element Method (FEM).

13.3 FEM

In the field of engineering design, many of our problems such as heat transfer process, system dynamics, fluid mechanics, solid mechanics, etc. have focused on developing mathematical models for physical systems and solving them with analytical methods. The mathematical formulation by and large is complex and not possible by analytical methods. Further, the analytical methods are most often limited to very simple geometries such as rectangles, spheres, etc. and entails concrete understanding of physics behind the real phenomena. The obvious choice is then to accept the use of suitable numerical techniques. An alternative method for solving mathematical models is the numerical method. With the rapid advancement of computers, numerical methods have become a sound and invaluable engineering tool. Numerical methods are remarkably important to solve extremely complex systems since the solutions are often achieved fairly quickly. One of the well-known numerical methods is the Finite Element Method (FEM). FEM has been in development for over 50 years. The FEM method is a very powerful tool for getting the numerical solution of a wide range of engineering problems. The fundamental principle is that an original physical body or structure is divided into smaller discrete elements of finite dimensions called as "Finite Elements". The body or structure is now considered as an assemblage of these *elements* connected at a finite number of joints called as "nodes". The attributes of the elements are formulated, configured and combined in order to obtain the properties of the entire body. Combining the equilibrium equation of each element derives the governing equations of equilibrium for the entire body. This in turn signifies that the continuity is ensured at each node. The essential boundary conditions are incorporated and the equations of equilibrium are subsequently solved to find the variables such as mass, volume, temperature distribution, strain energy, stress, force, displacement, velocity and acceleration depending on the applications.

Initially FEM was applied to the structural analysis of complex objects such as aircraft wings and calculations. But the method has been extended to other disciplines such as heat transfer, fluid flow and electromagnetics. The elements can model mechanics, acoustics, thermal fields, and electromagnetic fields. In an MEMS problem, the elements may model membranes, beams, thin film, plates, cantilever, thick plates, spring, mass, rigid element, fluids, etc.

In summary it can be stated that instead of solving the problem for the entire structure or body in one operation, the method advocates the formulation of properties of the constitutive representative elements. The method is thus modular in approach. The implied procedure essentially follows four

steps: formulate the element; combine the elements; get the solution of equations, and evaluation of the required variables. The complete system could be irregularly shaped and complex, but the constitutive elements are easy to analyse. The elements can be 1-dimensional, 2-dimensional (triangular or quadrilateral), or 3-dimensional (tetrahedral, hexahedral, etc.) (Fig. 13.2) and linear or higher-order. The behavior of a particular type of element is analysed in terms of the cause and effects (equivalently, loads and responses at discrete nodes, respectively).

The essence of the FEM method is to take a complex problem. Then it is decomposed into number of pieces upon each of which a simple approximation of the solution has to be made, and then the local approximate solutions are put together to obtain a global approximate solution. According to Michael A. Sprague (U. of Colorado) the above four steps can be expanded as follows.

Stpe-1: Identify the system governing equation

Step-2: Introduce an integral form equation

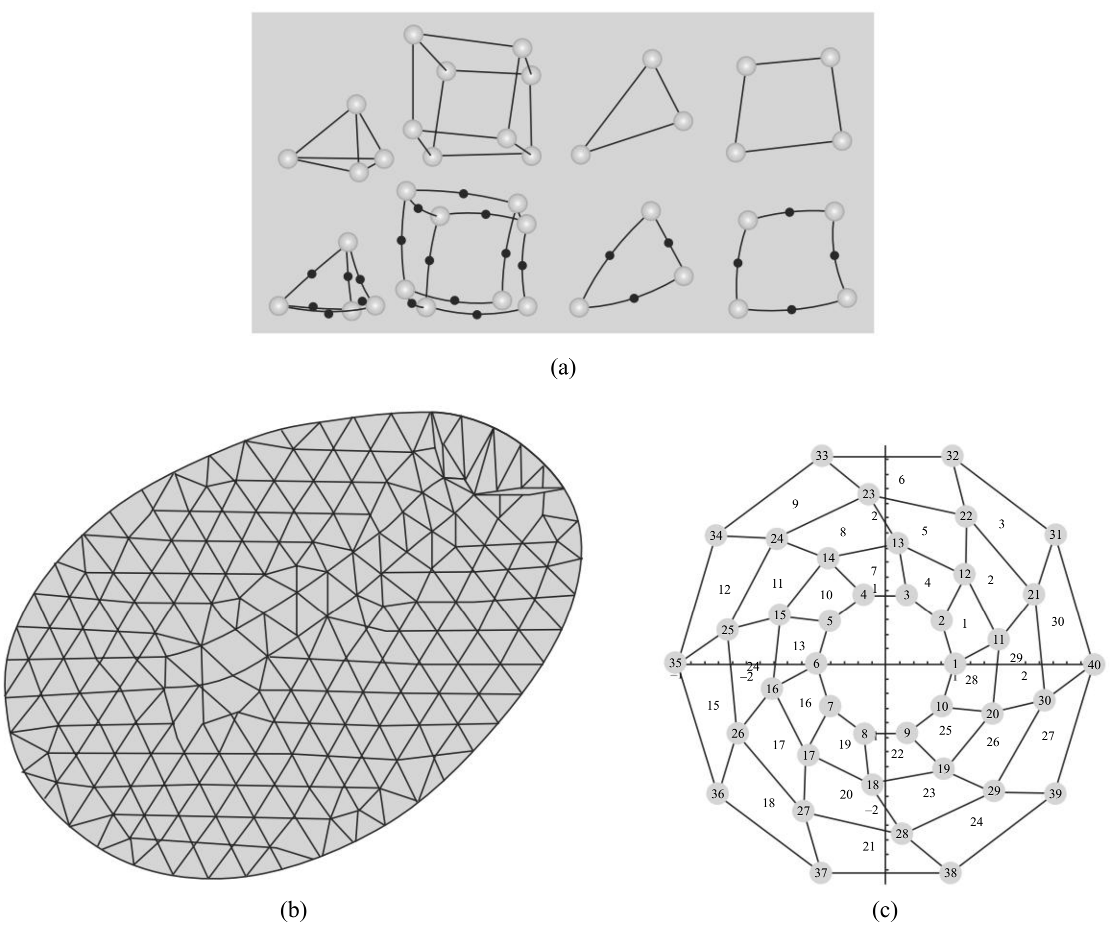


Fig. 13.2 *(a) Typical elements and the position of their nodes (Source:* R Funnel, Department of BioMedical Engineering McGill University*); (b) A typical ring-type structure, (c) Typical triangular elements of a portion within the structure (Source:* © Wolfram Research, Inc.*)*

Step-3: Descretize the domain of interest into elements
Step-4: Introduce an approximation of the field variable over an element
Step-5: Evaluate the integral form over each element
Step-6: Assemble the global matrix equation
Step-7: Solve the matrix equation to get the unknowns
Step-8: Calculate the values of interest from the approximate solution

13.4 DESIGN FLOW USING SIMULATION TOOL: EXAMPLE WITH MOEMS DEVICE

The MEMS devices are simulated within the GUI-based simulation environment. The GUI-based environment is called high-level system. The design flow[1] within the environment is a sequential process as shown in Fig. 13.3(a). Consider a MOEMS device. At first, the system specifications are incorporated to the integrated high-level system to foster the required MOEMS subsystem performance. The initial setting of specifications is mandatory in order to select a particular design and fabrication approach. The choices made at this stage include digital versus analog control, surface versus bulk micromachining, actuation method (thermal, electrostatic and magnetic), and the range of desired motion. The next step is to build a system *image*. The image, built from parametric primitives, is schematic in nature. The image is considered as a virtual description of the real device or system. At this stage graphical entities of electromechanical and optical components would be necessary. So there must be a provision of electromechanical and optical libraries within the high-level system. Importing respective components from the libraries and integrating them together create the image. The integration is called configuration. Figures 13.3(b) and (c) show some of the components in both the libraries. Coventor, Inc. has developed an excellent electromechanical and optical library, which can be used for the simulation of MOEMS devices. The libraries have been developed to allow rapid creation of 6-DOF (Degree-of-Freedom) electromechanical models for micromachined MEMS and MOEMS devices such as mirrors, resonators, accelerometers, gyros, and so on. A glimpse of a typical Gaussian beam in the optical library is shown in Fig. 13.3(d). Normally, the optical components are connected via a standard optical interface representing a Gaussian beam. As can be seen, eleven wires are necessary to define the beams characteristics as follows: beam power, wavelength, Rayleigh range (z_0), position of the previous component along beam z-axis (z_s), beam position in the reference frame both translational *x, y, z* and rotational *rho, theta, gamma* and finally the phase of beam at beam waist *phase*. Once the replica is satisfactorily completed, a device layout is then generated from this high level description. Detail descriptions of all the entities of various libraries are beyond the scope of this book. This section simply provides an example that suggests the design flow involved within the simulation environment.

13.5 ANSOFT DESIGNER™ AND HFSS V9.0

Ansoft corporation is pioneer in developing software products for electromagnetic simulation applications. Ansoft Designer™ and HFSS V9.0 are the two products, which can be used for simulating MEMS devices for RF applications. Ansoft Designer™ is a fully integrated high-frequency, physics-based electromagnetic simulation, modeling, and automation tool that is embedded into a seamless

[1] MEMS design flow and design model vary significantly. The information on MEMS design model is briefly presented in Appendix D.

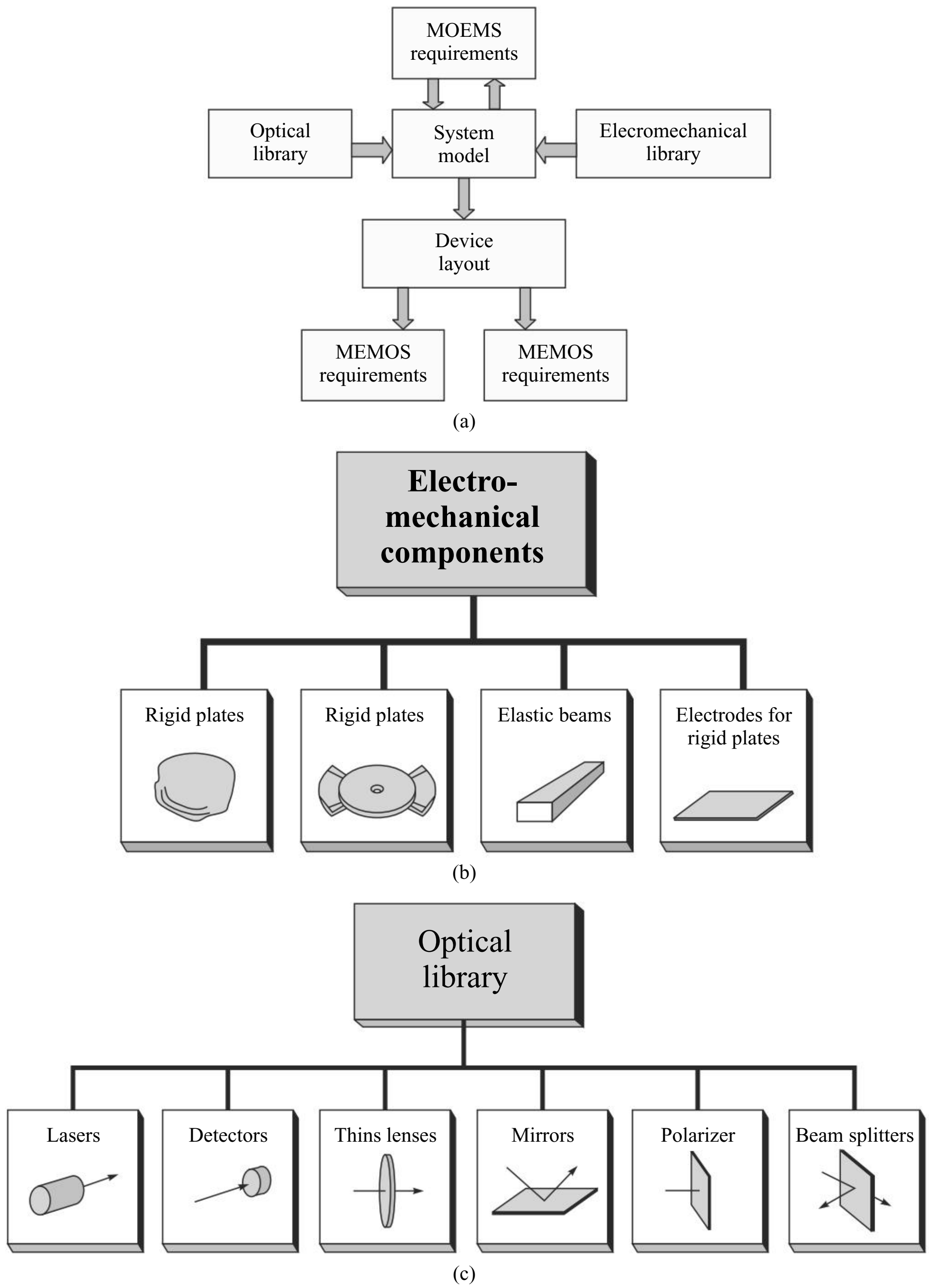


Fig. 13.3 *(a) Design flow of MOEMS devices within the simulation environment, (b) Electromechanical components for the MOEMS design, (c) Optical components for MOEMS design,*

Contd.

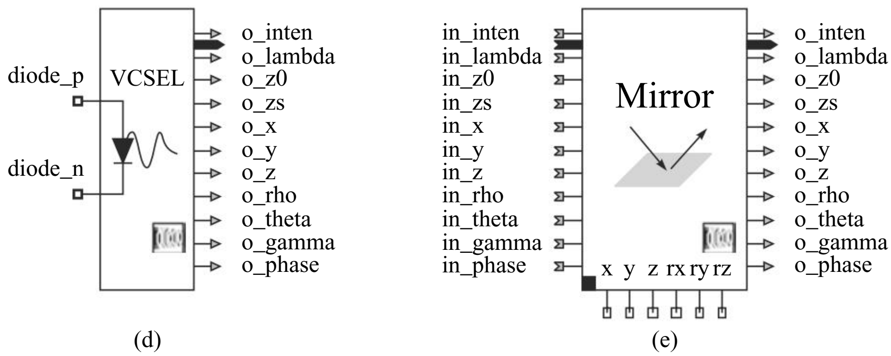


Fig. 13.3 (*Cond.*) *(d) The image of a Gaussian beam component of the optical library, (e) The use of Gaussian beam in building an image of a micromirror (Source:* Lorenz, Morris et al., Coventor, Inc.*)*

environment for circuit and system analysis. The design, fabrication and packaging of RF MEMS devices and optical integration of MEMS can all be analyzed and optimized by using these tools. Devices like Ka-band micromachined transmission lines, antennas, filters, resonators and switches can be simulated. The simulation environment incorporates developing strategy in terms of analysis of electrostatic, mechanical and electromagnetic behaviors of the individual and collective elements. Physical attributes and electrical performance are linked through Ansoft Designer kernel. The important features of the platform are given below.

- 3D electromagnetic structure simulator
- Digital-communication-system simulation can be performed
- Nonlinear circuit simulation
- Frequency domain and transient analyses
- Planar EM simulation
- Integrated IC and PCB layout editor
- Impulse invariance convolution engine
- Swept parameter analysis
- Real-time filter tuning
- Load-pull analysis
- Time, frequency, or mixed-mode domain analysis
- Generation and application of complex digital waveforms to arbitrary topologies

High-Frequency Structure Simulator (HFSS) is an electromagnetic simulator that is an integrated architecture with automation for enhanced design, analysis, and post-processing functionality. It incorporates optimetrics[2] into the systems. Optimetrics is a smart optimization engine that allows users to perform parametric analysis, optimization, sensitivity analysis, and many other design studies from an easy to use interface. Optimetrics can be applied to optimize the design of electromechanical products like motors, solenoids, and sensors; electronic products like printed circuit boards, IC packaging, and connectors; and high-frequency products like antennas, and microwave circuits. The designers can use parametrics to get close to the right solution, then optimize to get the maximum performance. Optimetrics allows users to:

[2] Optimetrics™ is trademark of Ansoft Corporation.

- Configure parametric controls on geometry shape and position
- Optimize material properties associated with the parts within a component
- Change boundary conditions to satisfy design criteria
- Utilize mathematical optimization methods to fine-tune the system

Optimetrics performs statistical analyses and sensitivity studies. Statistical analyses along with sensitivity studies provide insight into the performance of fabricated designs. This in turn gives benefits as far as automated design refinement and design-for-manufacturing approaches are concerned.

13.6 DS/MEMS AND CA/MEMS

3-D multiphysics analysis codes utilizing various numerical methods are the prime concern of the MEMS simulation tool. Note that microactuators and microsensors have multidisciplinary nature of operating principles. For instance a micropump is driven by a electromagnetic force (Lorenz force). Some devices use a thermal bimorph[3] effect for actuation. In order to handle such systems the need for multidisciplinary optimization (MDO) and multiphysics analysis tools for simultaneous observations are of paramount importance. KAIST (Korea Advanced Institute of Science and Technology) has developed a simulation platform for MEMS, which deals with the MDO. The name of the software is called Design System for MEMS or DS/MEMS in short. DS/MEMS is based on a CAD platform and window environment. The platform is simple to use, flexible and user-friendly. It enables a user to make a CAD model and formulate an optimal design problem within the windowed domain. It has been claimed that the performance and efficiency of DS/MEMS has been very satisfactory and the developed system is shown practicable for a multidisciplinary analysis and optimization of complex structures found in MEMS designs. DS/MEMS consists of two modules such as

- Optimal design module
- Robust design module

The CA/MEMS is the simulation engine for DS/MEMS, and thus act as the kernel. The CA/MEMS kernel includes various modules individually responsible for the analysis and optimization of the following sector.

- Structural
- Thermal
- Electric
- Electromagnetic
- Fluidic fields

It is also capable of analyzing various coupled-field problems. DS/MEMS handles interactions among programs automatically. The optimal design is performed by using performance characteristics extracted from the various analysis results. The optimal design module is composed of three optimization engines. They provide users with three methods such as nonlinear programming with finite difference method (FDM), Taguchi parameter design and the response surface method. The robust design module is used to minimize the perturbation of performances of MEMS devices under

[3] One of the electro-thermal actuation schemes is called bimorph effect where two materials with different thermal expansions are combined in a cantilever, known as bimorph cantilever. The thermal coefficient of the expansion mismatch between two layers of the cantilever can produce bending of the cantilever and thus vertical actuation if current is passed through the layers.

uncertainties. The example of uncertainties could be process tolerance and the change of operating environments. DS/MEMS has been tested through many MEMS structures such as micro-pump, micro-spatial-light-modulator, and micro-mirror that are operated by electromagnetic force, electrostatic force and thermal bimorph effect.

13.7 FEMPRO

FEMPRO is a simulation tool from ALGOR. The tool is based on finite element modeling and is useful for design, evaluation and optimization of MEMS devices. FEMPRO supports wide range of simulation capabilities which include,

- Static stress and Mechanical Event Simulation (MES)
- Linear and nonlinear material models
- Piezoelectric material models for MES
- Steady-state and transient heat transfer
- Steady and unsteady fluid flow
- Electrostatics
- Multi-physics analysis

FEMPRO also includes a suite of modeling and meshing tools and a wide range of result evaluation and presentation options. A built-in graphics environment provides extensive results evaluation and presentation capabilities, transparent display options, multiple-window displays, fast dynamic viewing controls and customization options including user-defined color palettes and annotations. All analysis results can be displayed graphically as contours or plots and the output can be displayed in the BMP, JPG, TIF, PNG, PCX and TGA formats. Figure 13.4 shows a snap shot of FEMPRO while electrostatic analysis software was used to calculate the electrostatic forces generated when voltage is applied to this MEMS radial comb motor.

13.8 ANSYS MULTIPHYSICS™

ANSYS software has evolved into a multiphysics multiengineering (MPME) analysis environment capable of simulating a broad range of physics, which includes,

- Structural
- Thermal
- Fluid
- Acoustics
- Electromagnetics
- Electronic circuits

The ANSYS Multiphysics tool supports direct and indirect coupling between different physics to facilitate accurate modeling of real world effects. Some of the well known MPME examples that are relevant to the MEMS domain are fluid-structural, thermo-electric, electro-structural (piezoelectric) and electrostatic-structural. Utilizing ANSYS software, one can simulate these phenomenon and their effects. ANSYS Multiphysics also integrates the power of matrix and sequential coupling to combine the appropriate "physical fields" required for accurate, reliable simulation results. The software can simulate complex thermal-mechanical, fluid-structural and electrostatic-structural interactions, and includes the complete range of powerful matrix solvers. In summary it can be stated that the simulation capabilities needed for complex design of MEMS are integrated in one environment that can provide comprehensive analysis and simulation results of structural, thermal, CFD, acoustic and

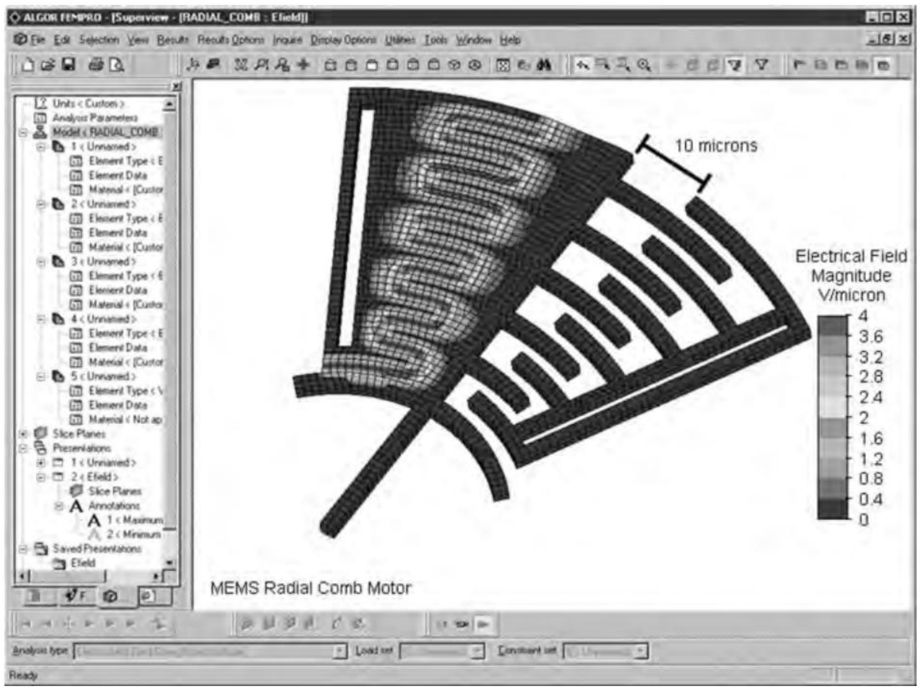


Fig. 13.4 *A snap shot of FEMPRO: used to calculate the electrostatic forces generated when voltage is applied to this MEMS radial comb motor (Courtesy:* ALGOR Inc*).*

electromagnetic attributes. The comprehensive features included in ANSYS Multiphysics are listed below. Table 13.1 shows the analysis capability relevant for a range of MEMS applications.

- Structural static, modal, harmonic, transient mechanical deformation.
- Large deformation structural nonlinearities.
- Full contact with friction and thermal contact.
- Linear and nonlinear materials (Buckling, creep).
- Material properties: Temperature dependent, isotropic, orthotropic, anisotropic.
- Plasticity, viscoplasticity, phase change.
- Electrostatics and magnetostatics (Low and high frequency electromagnetics).
- Circuit coupling—voltage and current driven.
- Acoustic—Structural coupling.
- Electrostatic—structural coupling.
- Capacitance and electrostatic force extraction.
- Fluid—Structural capability to evaluate damping effects on device response time.
- Microfluidics: Newtonian and non-Newtonian continuum flow
- Temperature dependent surface tension.
- Charged particle tracing in electrostatic and magnetostatic fields.
- Electro-thermal-structural coupling.
- Advanced thermoelectric effect such as Seebeck, Peltier and thermocouple.
- Piezoelectric and piezoresistive transducers: Direct coupled structural-electric physics.

Table 13.1 *ANSYS MEMS Applications (Courtesy:* ANSYS Inc.*)*

Microsystems	*ANSYS Multiphysics Capability*
Inertial Devices: Accelerometers and Gyroscopes	Structural modal, Static, Transient, Electrostatic-Structural, Reduced order macro modeling for system level.
Surface Acoustic Wave Devices	Acoustic—Structural coupling
MicroStripline Components	High Frequency electromagnetics.
Micro-patch and Fractal Antennas	High Frequency electromagnetics.
Piezo Inkjet Printheads	Thermal actuation: Electro-thermal—structural coupled physics. Thermal-structural coupled physics
Thermal Inkjet Printheads	Piezoelectric actuation: Direct coupled structural-electric physics. VOF Free surfaces & capillary action.
Micro mass spectrometers	Electromagnetics and charged particle tracing
Electrostatic comb drives	Electrostatic—structural coupling. Capacitance extraction.
Microfluidic Channels	Newtonian/non-Newtonian continuum flow
Piezoelectric actuators	Full isotropic and orthotropic parameters
Pressure transducers	Capacitance based: Electrostatic structural coupling. Piezo-resistive based: Electro-Structural indirect coupling
Electromechanical RF filters	Electrostatic—structural coupling. Capacitance extraction.
Micromirror technology	Electrostatic—structural coupling. Fluidic structural capability to evaluate damping effects
Micro-grippers	Electro-Thermal-structural
Micro TIP field emitters	Electrostatics and charged particle tracing
Micro-Gear assemblies	Mechanical with complex contact, friction.
Thermoelectric actuators	Electro-thermal—structural coupled physics
Magnetostrictive actuators	Low Frequency electromagnetics

13.9 SUGAR

SUGAR is a simulation platform, which has been developed by the Sensor and Actuator Center (SAC) at the University of California at Berkeley (UCB). SUGAR requires MATLAB release 5.2 or later and can be operated in Windows, Sun, HP and Alpha systems, and has been tested using MATLAB 6.0 on Linux systems. The code, demo files and manual are downloadable from the UCB. SUGAR inherits its name and philosophy from SPICE. A designer can describe a device in a compact netlist format, and very quickly simulate the device behavior. The main components of SUGAR are:

- A netlist interpreter
- Models describing the characteristics of the different components
- A command-line and GUI (Graphical User Interface)

The characteristics of the different components are written in MATLAB or C. The GUI allows interaction and visualization of the specified devices. The kernel of SUGAR handles nodes, elements and mesh assembly. MKS (meter-kilogram-second) system of units is used in SUGAR software. The standard notations are as follows.

d	deci	10^{-1}			
c	centi	10^{-2}	h	hecto	10^{2}
m	milli	10^{-3}	k	kilo	10^{3}
u	micro	10^{-6}	M	mega	10^{6}
n	nano	10^{-9}	G	giga	10^{9}
p	pico	10^{-12}	T	tera	10^{12}
f	femto	10^{-15}	P	peta	10^{15}
a	atto	10^{-18}	E	exa	10^{18}

Devices in SUGAR are described by input files called *netlists*. The basic unit of a SUGAR netlist is an element line. For instance,

crossbeam = element{A, B; model="beam2d", material=p1, l=100u, w=2u}

is an element line describing a beam. The following netlist can be created by opening a text editor.

```
-- 'cantilever.net'
use("mumps.net")
use("stdlib.net")

anchor {node "substrate"; material=p1, l=10u, w=10u, h=10u}
beam3d {node "substrate", node "tip"; material=p1, l=100u, w=2u, h=2u}
f3d    {node "tip"; F=2u, oz=90}
```

In order to allow users to experiment with variations on a simulation, variables are defined. Physical parameters associated with a particular layer of a particular material are called process parameters, and can be described as given below.

```
default = material {
   Poisson = 0.3,               --Poisson's Ratio = 0.3
   thermcond = 2.33,            --Thermal conductivity Si = 2.33e-6/C
   viscosity = 1.78e-5,         --Viscosity (of air) = 1.78e-5
   fluid = 2e-6,                --Between the device and the substrate.
   density = 2300,              --Material density = 2300 kg/m^3
   Youngsmodulus = 165e9,       --Young's modulus = 1.65e11 N/m^2
   permittivity = 8.854e-12,    --permittivity: C^2/(uN.um^2)=(C.s)^2/
                                                                  kg.um^3
   sheetresistance = 20         --Poly-Si sheet resistance [ohm/square]
}
```

The software supports three basic styles of analysis:

- Static analysis
- Linearized analysis
- Transient analysis

Static analysis attempts to find the equilibrium state for a MEMS device. The equilibrium state is characterized by a collection of force and moment balance equations. Linearized analysis is composed of modal analysis and steady-state analysis. Essentially, the behavior of the system near equilibrium is analyzed by approximating the system. Approximation to a system in the neighborhood of equilibrium provides valuable information on stability. Modal analysis studies the resonant behavior of the structure by solving the eigen problem. Here zero damping is assumed. Bode plot illustrating the amplitude gain and phase shift between a harmonic excitation at the input and a measured harmonic at the output can be produced through steady-state analysis. Through dynamic analysis, also called transient analysis, the response time, rise time and settling time can be studied. For more detailed explanation of SUGAR the readers are advised to refer SAC/UCB.

13.10 ATOMISTIC TO CONTINUUM THEORY

Material is discrete when viewed at an atomic scale and continuous when viewed at large length scales. Accordingly, we have atomistic and continuum models of the materials. Atomistic model express the material as *atom-by-atom* philosophy, whereas continuum model express the material in terms of *finite element* philosophy. The dual viewpoint of material is important in dealing with the physical phenomenon of the materials (components) at nano and microscale levels.

Atomistic model is sometimes referred to as Molecular Dynamic (MD) model. MD provides an atomic level picture of structure, which is simply the atomic property versus structure relationships. To characterize the collective behavior of the atoms in a material, acceptance of quantum mechanical descriptions of material is obvious. On the other hand, continuum model suggests that a material body is a continuous collection of a large number of deformable particles, each particle possessing finite size and inner structure. The model is also called continuum mechanics, which is eventually applied to describe larger material structure. In continuum mechanics the deformable particle is assumed to be *point mass*. The model reflects physical variables such as mass, temperature, stress, and so on, since the properties such as strength, tension, torsion, bending, etc. essentially depend on the size of the material structure.

Conventionally, the model is developed at a variety of *length scales.* Broadly the length scales are divided into five scales ranging from centimeter down to nanometer, as mentioned in Table 13.2.

Table 13.2 *Scale length and their range*

Length scale	*Range*	Continuum
Macro-scale	Millimeter to Centimeter	↑
Meso-scale	Hundreds of micrometer	
Micro-scale	Tens of micrometer	
Nano-scale	Nanometer	
Atomic-scale	Atomic and Molecular dimension	Atomistic (quantum)

There exist models corresponding to each length scale. Figure 13.5 shows various length scale and corresponding time scales with regards to various model theories starting from *ab initio* (Quantum) to continuum. The atomistic to continuum theory have been of prime importance in microengineering applications such as MEMS and NEMS design. The models are mostly useful in studying,

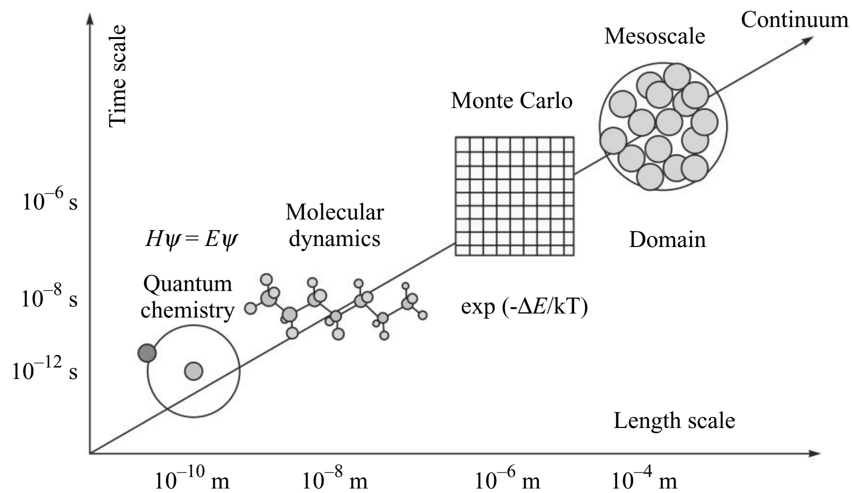


Fig. 13.5 *Length and corresponding time scales with regards to various model theories starting from ab initio (Quantum) to continuum.*

- Forces acting on particle-reinforced composites
- Torsion behaviour of extremely thin wires
- Loads on plates
- Bending in beams
- Indentation of a solid
- Melting, dynamics, phase transition phenomenon
- Crystallization
- Crack propagation and defects
- Diffusion and ion transport across membrane

13.11 TERMINOLOGY REVIEW

Before we present the various types of multiscale design methods, we like to define some of the important terms commonly used in this domain. For more detailed information readers are advised further reading on the topical subject *'micromechanics'*.

- *Mesh:* A functional region (part) in a MEMS or NEMS device.
- *Meshing:* Segmentation of the entire dimension of the structure/element into various functional regions and mathematically coupling them together through model equations.
- *FE region:* A region in a mesh, which can be treated under finite element method.
- *TB region:* TB stands for Tight Binding. A region in a mesh, which can be mediated or treated under Tight Binding method.
- *MD region:* A region in a mesh, which can be treated under molecular dynamics (MD) method.
- *Handshake region:* The overlapping regions in between FE/MD and MD/TB are known as handshake regions.
- *Scale:* Scale is a measure of length. The length can be high scale or low scale. The FE, TB, MD regions are reflected through scale.

- *Upscaling:* An approach to simulate the mesh from low scale atomistic level to high scale continuum level.
- *Downscaling:* The reverse of upscaling.
- *Dislocation:* A type of defect found in crystal lattices.

13.12 ANALYTICAL THEORY AND COMPUTATIONAL MODELING

The computational modeling of materials behavior has been a tool to strengthen scientific investigations and predictions to complement traditional, theoretical and experimental approaches. Efficient and accurate calculations of material behavior relies on adoption of simulation methods. Mechanical or physical phenomena such as crack growth (due to high loading or fatigue) dislocation and deformation all require explicit treatment of micro and nanoscale modeling details. To deal with the phenomenon a fundamental topical subject is the matter of study. The knowledge gained from this study can be utilized in dealing with many MEMS and NEMS designs.

The power of analytical theories lies in their ability to reduce the complex collective behavior of the basic component of a solid into insightful relationships between cause and effect. There is a requirement of describing the material phenomenological property by using a set of fundamental equations, called model equations, so that the behavior can be studied and predicted while designing the desired structural component for MEMS or NEMS devices. NEMS is a very recent development, which deals with the device design of the order of nanometer scale. The analysis will lead to the establishment of a sound technological podium for evaluation and optimization.

Although the set of governing equations, which can be used to describe the physical phenomenon of the material in various length scale mentioned above, are based on well-established physical laws but from their solution point of view they significantly differ as far as computation time and cost is concerned. The discrepancies in the solution occurs with regard to phenomenological properties such as,

- Time scale (Fig. 13.5)
- Strain fields
- Inelastic behavior
- Temperature
- Potential energy and energy density
- Resistivity/Conductivity, etc.

Continuum model analysis assumes that the energy density is spread smoothly throughout each element, but at the atomic scale, the potential energy is localized to the covalent bonds and the kinetic energy is localized largely to the nuclei. This small atomic scale mismatch can cause problems in finding the solution to calculating the potential energy, for instance. In another example, consider a material grain (a very small element) of size 5 μm^3 with 10^{13} atoms. For the calculation of resistance to deformation for instance, it is impossible to simulate this grain directly by using atom-by-atom philosophy. No such computational facility is available in order to the study such molecular dynamics and simulate the result. Note that the largest molecular dynamics simulation to date can handle 10^9 atoms (This is normally 1/10000th of the number of atoms in a real grain). It is therefore justified to accept the existence of the continuum model. The material can be modeled with a lower resolution using continuum techniques, called finite element method (FEM). Atom-by-atom philosophy is a higher

resolution level whereas FEM is a lower resolution level. Put it another way, the resolution in the continuum mechanics can be far coarser than in the molecular dynamics model. While the atomic resolution is required in very limited regions of space, in many situations the material can also be accurately modeled with lower resolution, such as FEM. However, neither one of these models can deal the solution to the microscopic level (Micro- and nanometer level) effectively due to the reason that the atomistic model is capable of handling the *atomistic region* (Also called molecular dynamics region, or simply MD region.) and the FEM can only deals with *continuum region*. An element in a micro device can be segmented into various regions called *dynamical regions.*

13.12.1 Pros and Cons

The design of MEMS and NEMS relies on meticulous understanding of the mechanics of the device element. A common difficulty with each domain is that an attempt to represent finer physical scales (high resolution atomic model) will lead to an enormous computational problem with long computation times and large memory requirements. The spatial (space, e.g. dimension) and temporal (time, e.g. simulation time) limitations of atomistic simulations leads to FEM model being considered. However, the discretization at a coarse level (low resolution FEM model) ignores the fine scale information resulting in a solution that does not provide meaningful elucidation, since in practice there involves many assumptions. As the element sizes get smaller, MEMS and NEMS are forced to operate in a regime where the assumptions of continuum philosophy is violated, and the accepted finite element models fail. The behavior of materials begins to be atomistic rather than continuous, giving rise to anomalous and often non-linear effects. For instance, if the design is performed by employing the FEM simulation method, the following consequence might occur.

"The element could become less stiff and more compliant than the simulation predicts and the dissipation principle is not adequate".

Therefore, the inadequacy of continuum model will become an obstacle to further miniaturization of MEMS, i.e. to the level of NEMS. In another example, for instance, the description of deformation beyond the elastic regime can be effectively modeled by appropriate constitutive equations, and the implementation of such relationships within continuum mechanics relies on the assumption that material properties vary continuously throughout the solid. However, certain phenomenological behavior linked to this deformation cannot be readily described within the framework provided by continuum mechanics. Still in another example, the dynamical regions smaller than 1 μm are significantly affected by atomic scale physics which in reality departs from the continuum elastic theory, and dynamical regions larger than 0.1 μm^3 exceed the current limit of about one hundred million atoms for atomistic simulation of silicon on a supercomputer, for instance.

13.13 MULTISCALE CONCEPT

The theory and modeling of complex physical systems either occurs at a single scale or widely separated scales with no interactions. The question will of which model to accept will now arise. Bear in mind that atomistic and continuum theories need and reinforce each other. This is where the multiscale methodology comes in play. The atomistic-continuum handshake is most effectively achieved within the framework of multiscale modeling. Multiscale methodology necessitates the development of link between the atomic and continuum scales, so that effective and efficient computational simulation can be achieved.

Figure 13.6, illustrates how the products and processes are developed and optimized through multiscale design methodology by the use of multiscale modeling approach and associated simulation platform. Multiscale concept is concerned with the new aspects of technologies and design paradigms that are required to optimally account for multiscale interactions in materials, devices and systems. The application of multiscale methodology will lead to the development of optimized products and processes.

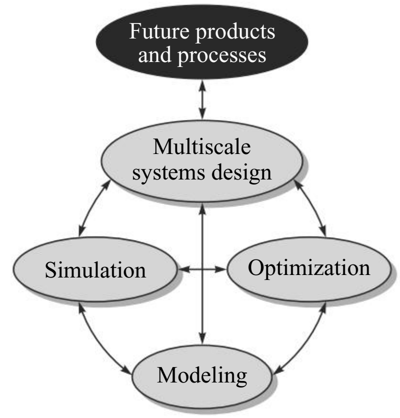


Fig. 13.6 *Optimization of products and processes through multiscale design methodology—an abstract level illustration (Source- Scientific Computation Research Center, Rensselaer Polytechnic Institute,)*

Many problems involve processes that occur over multiple length and time scales. Multiscale approach is considered as a concurrent method that couples molecular dynamics (MD) models, i.e. atomistic model, with the continuum models and is useful for studying many phenomenological properties of microsystems. It permits the use of far fewer equations than in strict molecular dynamics models. In these *coupled models*, the continuum domain serves as a boundary model. The multiscale modeling is considered as a 'divide and conquer' modeling paradigm, since the entire range of material behaviors is divided into a hierarchy of length scales as mentioned in Table 13.1. The relevant 'unit processes' (UP) are identified at each length scale. The UP processes at one-scale represent averages of unit processes operating at the immediately lower length scale.

13.13.1 Examples

Consider the example of a microresonator and a microgear system (shown in Fig. 13.7). Three regions having different scales are shown in this figure. With regard to the microresonator, the atomistic simulation (i.e. molecular dynamics; MD) is used in the regions with moderate strain oscillations. This constitutes one scale of the device. Finite elements (FE) is used in the peripheral regions, constituting another scale, where changes in the strain are small. The two are joined through a consistent boundary condition in the handshaking region (third scale). In case of microgear system, the structure has been decomposed into three regions and hence three scales. The inner region including the shaft can be treated by finite elements scale. This is due to this region's demand for the calculation or simulation of energy density that will drive the displacement field. The outer part of gears, excluding the regions at the gear–gear contact point, can be treated as molecular dynamic regions. The gear–gear contact point has been treated as tightly-binding (TB) region.

13.14 MULTISCALE METHODS

The ability to adapt advances in new design concepts to products will require a transformation in the methodologies of engineering modeling, simulation and design. The development of the multiscale design methodology for multiphysics systems will result in the dramatic improvement in the design technology. This in turn will affect the ultimate behavior of the entire system, and therefore the

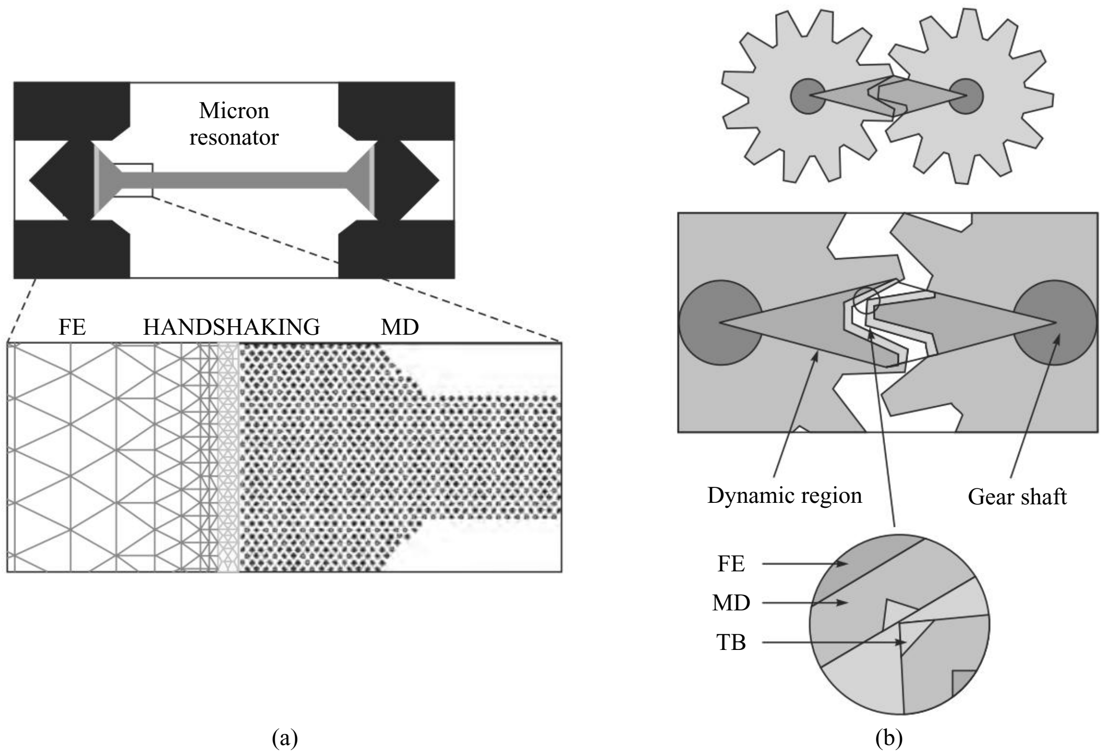


Fig. 13.7 *(a) Schematic diagram of typical length scales for a microresonator, (b) In case of a microgear train (Source:* Rudd et al., *Physical Rev.*, 58/10, 1998*).*

engineers must learn to model and design across this range of scales. Analysis of single scale and single physics is well developed. However, multiscale simulations will need to be enabled by multiscale mathematics. The physical systems involve highly nonlinear interactions among many phenomena at many different scales. The multiscale horizon entails deeper understanding of the phenomena from the viewpoint of how each scale interacts with each other, from the atomic scale through the microscopic to the macroscopic. Out of many developed multiscale methods the following methods are gaining momentum. They are,

- Macroscopic, atomistic, *ab initio* dynamics (MAAD)
- Quasi contiuum method (QCM)
- Bridging Domain Method (BDM)
- Coarse Grained Molecular Dynamics (CGMD)
- Coupled Atomistic Dislocation Dynamics (CADD)

Note that quantum and atomistic models and theories may not be able to simulate the continuum complexes. Simulation of large number of molecules is a time consuming task, no matter how many computer resources are thrown into the problem. Although computers are getting faster, but the researcher also wants to simulate further larger systems. One way to treat larger systems is to introduce simplifying approximations to the molecular theory in a logical way that the resulting model equations are simpler and faster to solve. It is thus essential to develop scale dependent theories, which should

bridge the gap between the continuum and atomistic model in order to determine and predict the material properties and mechanical response, respectively at the meso-, micro- and nanolevels. The simulation study will undoubtedly help in analyzing, studying and finally designing the micro/nano structures, such as MEMS and NEMS devices.

The physical laws governing the continuum property are reflected through three conservation attributes, namely mass, momentum and energy. Assuming adiabatic conditions and by adopting Lagrangian description, the conservation of mass and the conservation of energy are expressed as algebraic equation and ordinary differential equation, respectively.

13.14.1 MAAD

The unified macroscopic, atomistic, ab initio dynamics (MAAD) description brings all three descriptions together into a seamless union, embracing all the size scales, from the very big to the very small. The methodology adopts the concurrent linking of tight binding (TB), molecular dynamics (MD) and finite elements (FE) in a unified approach. Concurrent linking taken to mean as the simulation of these three regions can occur simultaneously. The approach suggests that the finite element is segmented until the size (technically called *mesh* size) becomes the order of the atomic spacing. At this level the dynamics are governed by the use of MD theory. TB is used to simulate the atomic bond breaking processes. The overlapping regions that are in between FE and MD as well as MD and TB are defined as "handshake" regions.

The MAAD scaling is downward, i.e. downscaling. The downward scaling has the advantage that the technique can eliminate spurious wave reflection at the FE-MD interface. However, the technique inherits disadvantages such as,

- Many timesteps are wasted
- The bulk assumption in constitutive relations disappears violating the constitutive relation

The equations of motion for FE, TB and MD are all integrated forward using the same timestep. The total energy can be calculated as,

$$H_{tot} = H_{md} + H_{md/tb} + H_{md/fe} + H_{tb} + H_{fe} \tag{13.1}$$

The above equation is self explanatory. The energy in the handshake regions is calculated based on the weight factors chosen, depending on which computational method contributes the most energy.

MAAD have been successfully tested in many applications including the simulation of brittle fracture. The brittle fracture simulation study requires spatial decomposition of the system into five regions. The FE scale region can study the continuum elastic energy which is a function of the displacement field, a continuous variable, that is integrated over the entire volume. This macroscopic description merely needs the constitutive law for the material. Around the microcrack, highly nonlinear deformation on the atomic scale is obvious. Further, there exists a large strain gradient around the crack. Molecular dynamic (MD) formalism is useful in such analysis. MD predicts the motion of the atoms governed by their mutual interatomic interactions and requires the numerical integration of Newton's classical equations of motion. In order to simulate and study the tip of the crack, where the failure of bond essentially occurs, tight-binding (TB) method is suitable. TB is considered as a semiempirical electronic structure description of matter. Through simulation study, the path of the crack can be tracked. Other two regions are the handshake regions exist between the FE, TB and MD regions.

13.14.2 QCM

The Quasicontinuum (QC) method is a mixed continuum and atomistic approach for simulating the mechanical response of materials (mesh). In the study, where the question like "Do We Need All of the Atoms for Simulation?" arises then the usefulness of quasicontinuum comes into picture. In large-scale molecular dynamics simulations, such as study of fracture, deformation, strain, strength, etc. in a particular mesh of the microdevice, the majority of the atoms within the mesh do not really participate in the process being studied. These atoms are sometimes referred to as *boring atoms*. In many simulations such boring atoms outnumber the atoms participating directly in the processes. It is therefore desirable to eliminate the boring atoms from the simulation process. By doing this, the computational time can be reduced since the approach does not consider all the atoms in the mesh.

Instead of treating all atoms, which make up the whole system, a small but relevant subset of atoms is chosen by appropriate weighting to represent the entire system. The representative of atoms, called *repatoms*, may vary from one region to another depending upon the problem at hand. For instance, in the study of deformation problem, representation of atoms is of varying density with more atoms being considered in highly deformed regions and fewer in the less deformed regions. The density can be adaptively updated as the deformation evolves. The inherent theory behind the QCM method is that it systematically coarsens an atomistic description by introducing and utilizing the kinematic constraints judiciously. Note that QCM considers the material behavior by inputing atomistic theory and adopting approximations that are strictly kinematic in nature.

For energy calculation, the theory starts from an underlying atomistic model, which delivers the energy of the crystal as a function of the atomic positions. Identifying the subset, a representative of atoms, can then reduce the spatial parameter of the mesh. This subset behaves as the independent degrees of freedom of the mesh. The positions of the remaining atoms are obtained by piecewise linear interpolation of the representative atom coordinates. QCM consists of an optimization of the energy for which no time or temperature phenomena are present.

Let us examine displacement field when subjected to deformation. Let us assume that the reference state of the body is X. A body deforms from reference state to the deformed state $\boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}(\boldsymbol{X})$. It should be pointed out that the key measure of a displacement field is the deformation gradient $\boldsymbol{F}$, which is given by

$$F(X) \equiv \frac{\partial x}{\partial X} = I + \frac{\partial x}{\partial X} \tag{13.2}$$

where, $\boldsymbol{I}$ is the identity tensor. Bearing in mind that the deformation gradient changes gradually on the atomic scale, we have to only track the displacement of a small fraction of the atoms explicitly, whereby the displacements of the remaining atoms are approximated through interpolation. This consideration is called reduction of degrees of freedom. The total energy is expressed as,

$$E^{tot,h} = \sum_{i=1}^{N} E_i(u^h) \tag{13.3}$$

where, $\boldsymbol{u}^h$ is the atomic displacements, implicitly found through the interpolation functions such that,

$$u^h = \sum_{\alpha=1}^{N_{rep}} S_\alpha u_\alpha \tag{13.4}$$

where S_α is the interpolation function associated with repatom α, and N_{rep} is the number of repatoms ($N_{rep} << N$).

13.14.3 BDM

In bridging domain method the continuum and molecular domains are overlapped in a bridging subdomain, where the Hamiltonian[4] is considered as linear combination of the continuum and molecular Hamiltonians. The atomistic and continuum model overlap at their junctions in a bridging domain. This method has the following advantages,

- Eliminate spurious wave reflection without any additional filtering or damping
- Projects the fine scale solution onto the coarse scale solution in the bridging domain
- Filters out the high frequency components at the interface
- Uniform mesh can be used in continuum subdomain
- Multiple timesteps can be implemented into this method for different length scales
- The method is not based on linearization so it can be applied to nonlinear problems

The method provides a natural way to couple heat conduction and other diffusion phenomena between molecular and continuum models.

13.14.4 CGMD

Coarse-grained molecular dynamics (CGMD) approach is somewhat related to the MAAD approach. This approach removes the TB method from the MAAD method and instead couples only FE and MD. This proposal claims that it is convenient to distinguish between two regions of the simulation, effectively. Thus, primarily there exist only two regions. The region in which the nodes are in one-to-one correspondence with the atoms is called the molecular dynamics (MD) region and the region with many atoms per node is called the coarse-grained (CG) region. The explicit idea in CGMD is that a coarse-grained energy approximation, which converges to the exact atomic energy, is utilized to derive the governing equations of motion. The equation of motion for the atoms in the MD region is similar to equation of motion based on empirical potential. In the CG region, the equation of motion is of the finite element form. The coarse grained energy from which the equations of motion are extracted is expressed as,

$$E(u_k,\ \dot{u}_k) = U_{\text{int}} + \frac{1}{2}\sum_{j,k}(M_{jk}\dot{u}_j \cdot \dot{u}_k + u_j \cdot K_{jk} \cdot u_k) \tag{13.5}$$

where,

Internal energy	U_{int} given by $3(N - N_{node})kT$
Kinetic energy	$M_{jk} \cdot u_j \cdot u_k$
Potential energy	$u_j \cdot K_{jk} \cdot u_k$
Stiffness matrix	K_{jk}
Mass matrix	M_{jk}
Displacements	u
Velocities	$\dot{u}$

[4] The value of the Hamiltonian can be interpreted as the energy of the system.

The stiffness and mass matrix are obtained using weight functions, which are similar in form to finite element shape functions. The internal energy is the thermal energy of those degrees of freedom, which have been eliminated, i.e. coarse grained out of the system. The equation of motion is expressed as,

$$M_{ij}\ddot{u}_j = -G_{ik}^{-1}u_k + \int_{-\infty}^{t} \eta_{ik}(t-\tau)\dot{u}_k(\tau)d\tau + F_i(t) \tag{13.6}$$

where, M_{ij} is the mass matrix, G_{jk} is a stiffness like variable, η_{ik} is known as memory function and $F_i(t)$ is a random force.

Coarse-grained molecular dynamics has many advantageous points such as

- Permitting large volumes of the system to be simulated with a reduced number of degrees of freedom
- CGMD automatically gives the MD equation of motion, as the cell size is reduced to the atomic scale
- Higher quality interface between the MD and CG regions
- Unwanted elastic wave scatter is reduced significantly
- Thermal expansion effects are treated properly

However, CGMD is more expensive computationally than conventional finite element modeling.

13.14.5 Coupled Atomistic Dislocation Dynamics (CADD)

To handle multiple scales simultaneously, there exists another method, which is called Coupled Atomistic and Dislocation Dynamics (CADD) method. CADD is introduced wherein atomistic and continuum regions communicate across a coherent boundary. In 2D models it facilitates exchange of dislocations back and forth as dictated by the mechanics of the problem. The atomistic region can experience any deformations that occur under the applied loading while the continuum region evolves according to dislocation plasticity.

13.15 COMPLEXITY OF MULTISCALE SYSTEMS

Multiscale Simulations covers atomic-nano to meso to continuum-macro scale phenomena. It is considered as a multifidelity simulations environment using analytical, behavioral or reduced models. Most MEMS and NEMS entail a very diverse range of length and time scales, and require a hierarchy of modeling approaches. Analysis and observation of mechanical response of microstructures through simulation by utilizing multiscale methodology covers many scales of resolution. The simulation involves many processes. Each process that is attempted to model and simulate must be considered in the context of its own spatial and temporal (time scale) resolution. If the model involves a single process, the scales can be set accordingly. This is not a serious problem. However, if multiple processes are active in a simulation, as in case of multiscale technique, the decision of what scale to use can affect either the accuracy of the models or result in severe penalties for computational efficiency. The success lies with the thorough understanding of the both temporal and spatial characteristics of the structure at hand. This in turn requires some fundamental adjustments in the way modeling is approached.

Coupling molecular level simulation techniques with continuum level is challenging but critical for successful designs. At the continuum scale, information such as velocities and pressure are relatively smooth functions of spatial and temporal variables with little stochastic character. On the other hand,

nanoscale simulations generate information that can only be characterized in terms of a time or ensemble averaging of dynamic variables in order to equate them with a continuum variable such as temperature and pressure. The challenge to computational models is the range of significant time scales ranging from femtosecond for rapid atomic motions to microsecond for experimental observations such as flux, for instance. Techniques that can capture all time scales in explicit detail within the scope of a single simulation would require a tremendous amount of computer time.

13.16 MULTIPHYSICS-MULTIENGINEERING INTEGRATION: AN ILLUSTRATION

Multiphysics-Multiengineering (MPME) integration looks at systems design that adheres to the basic principles of macroscale and nanoscale integration methods considering physics, chemistry, biology and engineering in terms of exploring issues like scalability, modularity, interoperability, integrability. The methodology embeds computing methods, simulation and displaying of optimized design parameters at all length scales, from the nanoscale to microscale to macroscale in a GUI (Graphical User Interface) environment.

Figure 13.8 expresses the vision of MPME multiscale integration philosophy, typical to a microfluidic MEMS and NEMS design. The design methodology integrates fundamentals, principles, theories and concepts. Such systems are composed of microchannels or microwells, which handle fluids of the order of nanoliters or picoliters. An example of microfluidic system has been taken due to the fact that the microfluidic technologies are emerging and providing opportunity to participate in the multiscale revolution. These microfluidic devices incorporate many of the transduction concepts, design criteria and performance parameters.

13.17 IMPORTANT FEATURES OF CAD TOOL

Computer Aided Design (CAD) tools for MEMS are being developed for electronic circuit designers as well as the MEMS experts. In MEMS domain, CAD is defined as a flexibly organized set of cooperating software modules that enable the simulation of implementation processes, device operation and microcomponent behavior in an iterative sequence. One of the important characteristics is that the tool itself should be a hierarchy of representations and models of MEMS to satisfy the needs of a spectrum of end users. Figure 13.9 illustrates MEMS design practices, which focus on iterative device and process development. The design practices entail parallel views and supporting libraries in the MEMS design and synthesis methodology. Three parallel views such as physical, structural and behavioral are shown in the figure. As always, the design starts with the specification and concepts. The physical view always deals with the 2D layout and the 3D model. The structural view identifies the connectivity between the elements in a typical schematic. Finally the behavioral view specifies the governing fundamental laws and equations of the system. This includes electrical node equations, heat transfer equations, motion equations, and so on.

Visualization of 3-D structures that result from a given layout and process is considered the biggest obstacle for MEMS designers. The implementation of the following features helps facilitating design and virtual prototyping of MEMS. Important features of a versatile MEMS CAD tools should include:

- An integrated toolset with capability to design at different levels
- Generation and extraction of different views

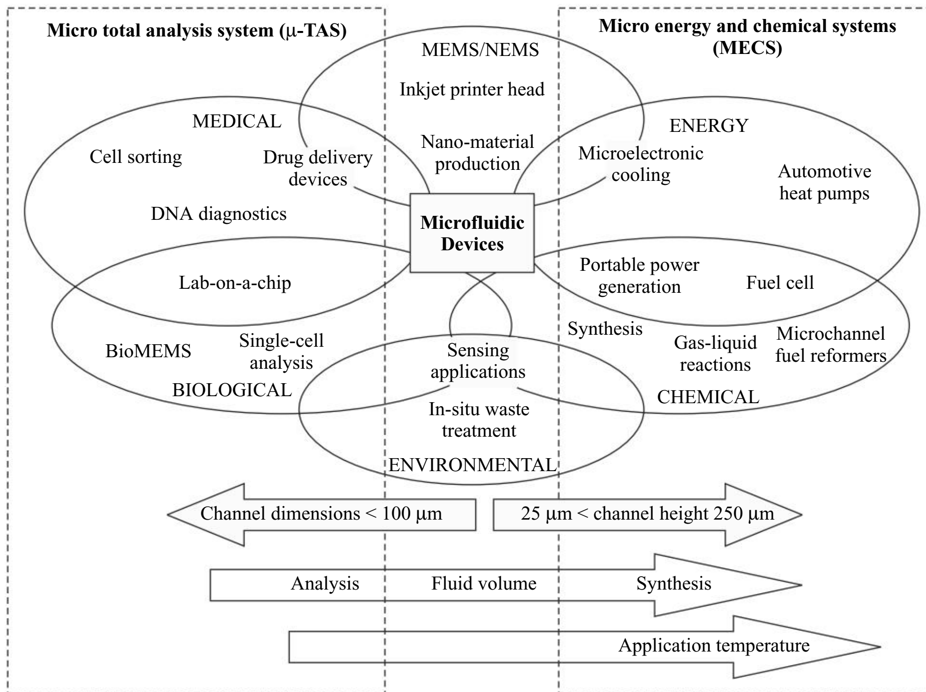


Fig. 13.8 *Multi-scale systems integration of; An illustration with microfluidic device (Source:* Paul, Springer, 2005*)*

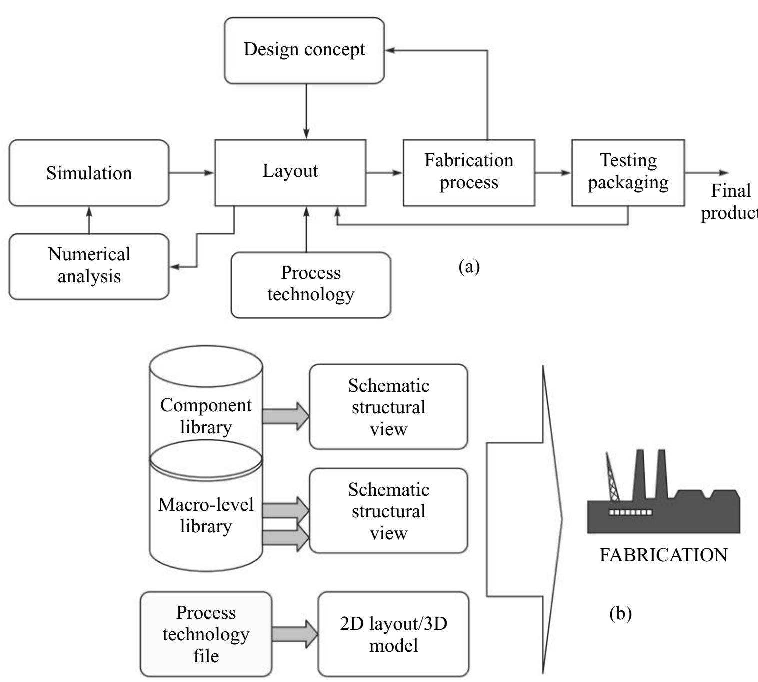


Fig. 13.9 *MEMS design practices*

- Fast and accurate algorithms for multi-domain engineering analysis
- Good numbers of potential libraries of reusable elements
- A layout and process representation that allows efficient representation of
    - Arcs
    - Freeform shapes along
    - Cross-sectional information
    - Material properties
    - Mesh information.
- An integral layout, solid modeling, and viewing tool for visualizing 3-D structures that may result from the combination of layout and process description
- Data translators for sharing design information between existing and other MEMS specific tools.
- Mesh generation and analysis tools
- Simulators for computationally efficient modeling of MEMS at component, block and system levels
- Modules for synthesis of mechanisms
- Support for top-down and bottom-up design procedures

As many design issues govern the CAD representation of MEMS and VLSI frameworks, in order to see if a VLSI framework methodology is extensible to the MEMS field, the differences and similarities between the two must be explored in first place. Note that within VLSI domain, the major part of the design occurs in a single primary energy domain with interactions described primarily by electrical quantities, such as current and voltage. On the other hand, the MEMS components of the system work in multiple energy domains such as electrical, mechanical, thermal, chemical, magnetic, acoustic and fluidic. Further, there is a coupling between them. As a mater of fact the developed CAD framework for MEMS has to support representation and analysis in multiple energy domains. Moreover, in VLSI, a 2D layout abstraction is supported while in MEMS a 3D representation of structures is needed. A typical CAD tool featuring various aspects of design, analysis and interfacing domain is illustrated in Fig. 13.10.

13.18 SUMMARY

Simulation software are application programs that run as graphic-like platform for the conceptual design, interactive design, visualization, and evaluation without tying up the real design of physical processes. Simulation thus can provide insights into the designs of processes, architectures, properties, effects, before significant time and cost has been invested, and can be of great benefit. The virtual design environment can greatly improve reliability and availability and can help to shorten the design-to-implementation cycle by enabling the users to correct errors and to identify optimized design parameters and requirements.

In this context this chapter introduces the need for simulation tools for the design of MEMS components and devices. The main problem encountered in simulation is, however, scale dependency. Simulation study for optimal design of micro and nanosystems entail consideration of multiple length scales. Effective development requires the synergism of advanced computer-aided design (CAD), multi-physics computer-aided analysis and engineering, materials science and technology, and also of effective quantitative testing methodologies for characterizing the performance, reliability, and integrity of MEMS and NEMS components at the different levels of their molecular packaging.

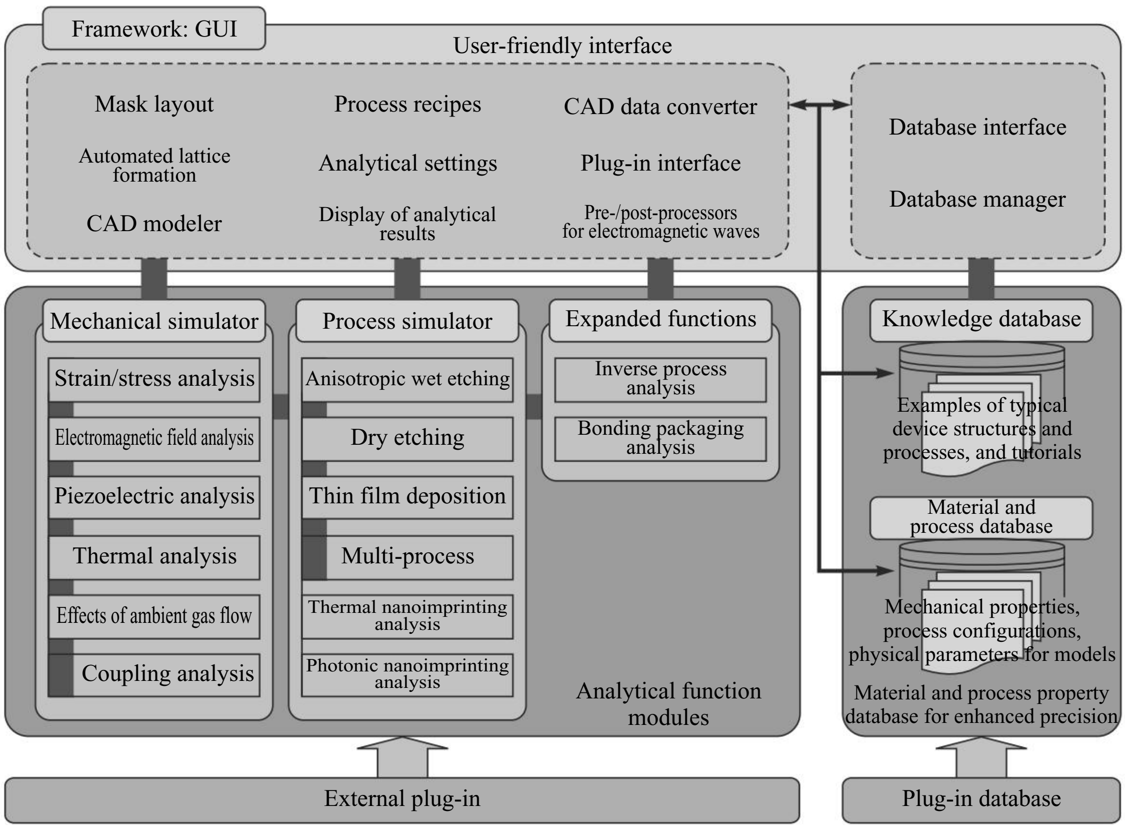


Fig. 13.10 *Important features of a versatile MEMS CAD tools*

Multiscale design approach is a new modeling technique that can deal with the design and simulation efficiently in terms of solving the geometric complexity and the scale dependent effects that usually arise. Continuum descriptions of micro and nano system become invalid as length scales decrease. Thus, methods based on molecular (atomistic, or quantum chemistry) are necessary. Moreover, non-continuum effects are often localized to small regions of large systems, preventing the global application of detailed models due to computational expense. Multiscale approach seeks to couple efficient continuum theory with detailed models to provide accurate and efficient models of complete systems. The multiscale modeling theory and corresponding simulation tool can facilitate the exploration of nanoscale phenomena with the goal of developing novel micro- and nanoscale devices.

This chapter describes the concept of multiscale design approach and simulation. Some of the available simulation software conforming to multiscale design theory are presented. In particular, the simulation software such as Ansoft designer, HSSS, DS/EMS-CA/FEMPRO, Ansys and SUGAR are introduced. Introduction to several multiscale design methods such as, Macroscopic, atomistic, *ab initio* dynamics (MAAD), Quasicontiuum method (QCM), Bridging Domain Method (BDM) , Coarse Grained Molecular Dynamics (CGMD) and Coupled Atomistic Dislocation Dynamics (CADD) are presented.

Points to Remember

- The development of new devices and their applications depends on having adequate CAD (Computer Aided/Assisted Design) tools for simulation and evaluation.
- Simulation is a design methodology, using which the designers study the performance of the system (an MEMS structure or component, for example) prior to its real design.
- In multiscale design, the objective is to predict the performance and behavior of structured materials across all relevant length and time scales utilizing fundamental physical principles.
- Atom-by-atom understanding of matter is reflected through atomistic model theory. On the other hand continuum theories describe a system in terms of some physical variables such as mass, for instance.
- The atomistic models and theories may not be able to simulate the continuum complexes due to the fact that simulation of large number of atoms is a time consuming task, no matter how many computer resources one throws into the problem.
- It is essential to develop scale dependent theories, which should bridge the gap between the atomistic and continuum model in order to determine and predict the material properties and mechanical response, respectively at all the scale such as meso, micro, and even nano scales. There comes the idea of multiscale design philosophy.
- In computing environment, simulation software are application programs run as graphic-like platforms for the interactive design, programming and evaluation of systems.
- The design of any product can benefit from computer-assisted simulation and optimization tools prior to its prototype testing followed by real design.
- The mathematical formulation is by and large complex and not possible by analytical methods. An alternative method for solving mathematical models is the numerical method.
- The fundamental principle of FEM method is that an original physical body or structure is divided into smaller discrete elements of finite dimensions called as "Finite Elements". The body or structure is now considered as an assemblage of these elements connected at a finite number of joints called as "nodes".
- Instead of solving the problem for the entire structure or body in one operation, the FEM method advocates the formulation of properties of the constitutive representative elements.
- Some of the commonly used MEMS simulation software are HFSS, DS and CS MEMS, FEMPRO, ANSYS and SUGAR.
- Material is discrete when viewed at an atomic scale and continuous when viewed at large length scales.
- Atomistic model is sometimes referred to as Molecular Dynamic (MD) model.
- Atomistic model expresses the material as *atom-by-atom* philosophy, whereas continuum model expresses the material in terms of *finite element* philosophy.
- The theory and modeling of complex physical systems either occur at a single scale or widely separated scales with no interactions. Multiscale methodology necessitates the development of link between the atomic and continuum scales, so that computational simulation can be achieved.
- Out of many developed multiscale methods, Macroscopic, atomistic, *ab initio* dynamics (MAAD), Quasicontiuum method (QCM), Bridging Domain Method (BDM), Coarse Grained Molecular Dynamics (CGMD) and Coupled Atomistic Dislocation Dynamics (CADD) methods are gaining momentum.
- MAAD adopts the concurrent linking of tight binding (TB), molecular dynamics (MD) and finite elements (FE) in a unified approach.
- The Quasicontinuum (QC) method is a mixed continuum and atomistic approach for simulating the mechanical response of materials (mesh). Instead of treating all atoms, which make up the whole system, a small but relevant subset of atoms is chosen by appropriate weighting, to represent the entire system.
- In bridging domain method the continuum and molecular domains are overlapped in a bridging subdomain, where the Hamiltonian is considered as linear combination of the continuum and molecular Hamiltonians

(The value of the Hamiltonian can be interpreted as the energy of the system). The atomistic and continuum model overlap at their junctions in a bridging domain.

- Coarse-grained molecular dynamics (CGMD) approach is somehow related to the MAAD approach. This approach removes the TB method from the MAAD method and instead couples only FE and MD. This proposal claims that it is convenient to distinguish between two regions of the simulation, effectively.
- CADD is introduced wherein atomistic and continuum regions communicate across a coherent boundary. In 2D models it facilitates exchange of dislocations back and forth as dictated by the mechanics of the problem.
- Coupling molecular level simulation techniques with continuum level is challenging but critical for successful designs. At the continuum scale, information such as velocities and pressure are relatively smooth functions of spatial and temporal variables with little stochastic character. On the other hand, nanoscale simulations generate information that can only be characterized in terms of a time or ensemble averaging of dynamic variables in order to equate them with a continuum variable such as temperature and pressure.
- Multiphysics–Multiengineering integration looks at systems design and adheres to the basic principles of macroscale and nanoscale integration methods considering physics, chemistry, biology, engineering, in terms of exploring issues like scalability, modularity, interoperability, integrability.

Exercises

1. Simulation is a design process–Justify. Illustrate traditional versus simulation oriented design approach.
2. Why is FEM important in simulation domain? What are the underlying steps necessary to advocate FEM approach?
3. Describe the design flow within the simulation environment.
4. Write notes on the following.
   (a) Ansoft Designer and HFSS
   (b) DS/MEMS and CA/MEMS
   (c) FEMPRO
   (d) ANSYS Multiphysics
   (e) SUGAR
5. What do you mean by multiscale design approach?
6. Distinguish between atomistic and continuum theory/model.
7. Define the following terms which are frequently used in the multiscale design domain.
   (a) Scale
   (b) Mesh and Meshing
   (c) FE, TB and MD region
   (d) Handshake region
   (e) Upscaling
   (f) Downscaling
   (g) Dislocation
8. How is material behavior studied through computational modeling? Discuss the pros and cons of different analytical models.
9. Give an illustration of a multiscale design concept.

10. Discuss various types of multiscale design methods you know. In particular, briefly describe the following methods.
    (a) MAAD
    (b) QCM
    (c) BDM
    (d) CGMD
    (e) CADD
11. What are the complexities involved in multiscale design approach?
12. Give an illustration of a multiphysics multi-engineering multiscale design philosophy.

Chapter 14

Performance Indices and Device Array

Objectives

The objective of this chapter is to study the following.

- The performance parameters such as sensitivity, repeatability, resolution, accuracy, error and nonlinearity
- Transient response specification
- Hysteresis
- Stability
- Reliability and availability
- Traceable calibration
- Array devices and their advantages

14.1 INTRODUCTION

The quality of a device, equipment or system is assessed through many ways. The quality characterizes a particular aspect, capability or attribute of a system, which are separately quantified by numerical values through respective definitive terms. These terms are called performance indices. Various performance indices for measuring the quality have been defined. As is obvious, the performance indices are used to compare the performance of different devices, equipment or systems. This chapter describes various performance indices. Performance indices are also called performance-measuring parameters.

The second part of the chapter deals with array devices. When multiple and similar types of elements or components are fabricated in one platform in order to accomplish the same nature of work the device is called array device. The role of array device is two fold. In one case, comparing and analyzing the effects from several elements/components the errors can be minimized and hence the performance can be improved. Examples are microthermovessel arrays, sensor fusion devices, and so on. In other case, many systems encapsulate array configuration as a part of their design methodology. Examples are

DMD (Digital Micromirror Devices), GLV (Grating Light valve), FSS (Frequency Selective Surfaces), and so on.

14.2 PERFORMANCE PARAMETERS

Performance is perhaps more difficult to define but frequently understood as a particular set of expected effects that may fulfil the envisaged outcome. The manner in which a system operates is measured by the use of some parameters. In other words, the degree to which a device can function according to specific criteria and achieve results in agreement with stated goals is reflected through some dedicated parameters. These parameters explicitly imply a measure of how well or badly a device or system can function in known environments. The parameters are known as performance measuring parameters or simply performance parameters, as they are usually observable and measurable. The commonly encountered performance parameters are described below, however other important indices will be presented in sequel.

- Repeatability
- Resolution
- Sensitivity
- Accuracy and precision
- Nonlinearity

14.2.1 Repeatability

Repeatability is the ability to perform the given function iteratively. It characterizes the closeness of agreement between independent results obtained after short intervals of time, with the same technique or method on identical devices under the same conditions. This performance parameter although applicable to many types of systems and equipments, however, is frequently used to define the field devices such as sensor and actuator performance. It is very common to say that an actuator might possess poor repeatability due to loose encoders, tight sideways and/or accidental damage. As an example, although not quite friendly to MEMS realm, the repeatability is the ability of a single pitch, on successive rotations of the belt in a conveyor system, to return to a home position within a specified tolerance. This in turn attempts to elucidate that the repeatability is the extent to which successive attempts to move to a specific location vary in position. A highly repeatable system, which may or may not be accurate, exhibits very low dispersion in repeated attempts to a given position, in a positioning system for instance, regardless of the direction from which the point was approached. Quantitatively, it is defined as a ratio of maximum deviation to the full range of the system and usually expressed in percentage.

14.2.2 Resolution

Resolution describes the degree to which a change can be observed and detected. It is expressed as a fraction of a total to which a process can endure without introducing unbearable error. It is the smallest change in the measurand or the observation that causes a detectable change in the measurement.

Resolution is a design characteristic and reflects a variety of ways depending upon the type of application. In case of actuator, the resolution typically is determined by the resolution of the motion of the actuator. As an example, it can be said that 'piezoelectric actuators provide nanometer resolution compared to other types of actuator'. Other examples are given below.

- If it is required to produce a digital *word* with 8 bits, then the number of *quantization* levels should be $2^8 = 256$. Consider a transducer output (analog signal), which produces an electrical signal

within the limit of 0.05 volts to 3.0 volts. The range of the transducer is thus 2.95 volts. If you adopt the analog to digital (AD) process with a 4-bit word length, then the resolution would be $2.95/2^4 = 184$ mV. In this case 184 mV is the smallest charge that can be observed and detected. On the other hand, if you adopt an AD process with 8 bits of word length, then the resolution would be $2.95/2^8 = 11$ mV. While reconstructing, higher the number of bits, more accurately an analog signal can be reconstructed from its digital representation.

- The number of pixels in a digital display system is commonly referred to as its viewing resolution. Pixel is the smallest unit of an image. Displays are made up of hundreds of thousands or millions of tiny squares called picture elements, or just pixels. The number of pixels displayed per unit length of an image typically measures in pixels per inch; sometimes dots per inch (dpi). In describing computer screens, printers and scanners we use a resolution parameter, e.g. dpi. The monitor might display about 70–100 dots of information in every inch of screen (70–100 dpi). For better resolution one can select 90 dpi resolution set. The scanner can be set to scan the page with specified resolution. Similarly, a printer can be set to print at 600 dpi resolution. Besides its theoretical definition, the resolution of a system depends on the underlying technology the system has adopted. In the context of a state-of-the-art printer for example, presently the inkjet print heads are based on MEMS technology. Compared to conventional technology, the MEMS based print head has the potential to improve the quality of inkjet images by permitting larger and denser arrays of microink orifices of various dimensions to provide improved resolutions[1].
- In contrast to conventional display technology that uses cathode ray tubes and liquid crystal displays, the MEMS based GLV (Grating Light Valve) and Digital Light Processing (DLP) technologies are now providing high-resolution capability.

14.2.3 Sensitivity

Sensitivity is defined as the responsiveness of variable of interest to an independent forcing variable. Quantitatively it is defined as the ratio of full-scale output to the rated capacity of a given transducing or an actuating system. Full-scale output is sometimes referred to as span. The span is the difference between the output at the rated capacity and output at zero applied forcing function. The rated capacity is the maximum forcing variable a system is designed to measure (in case of sensor) or produce (in case of actuator). Sensitivity, primarily depends on the following four factors,

- Geometry of construction
- The materials used
- Internal perturbations
- Environmental impacts

14.2.4 Error

The understanding of error is very common in our everyday life. Error may occur in sensors, actuators and devices. Within a sensor typically the error is considered as the difference between the values provided by the transducer and the true value of the measurand that is being sensed. Similarly, in a servo-controlled actuating system the error is the difference between the output state and the commanded state.

[1] Note that the MEMS based orifice structures control the motion of ink by exploiting its surface properties. Heating a fluid meniscus non-uniformly induces a gradient in surface tension. The surface tension produces a tangential force on the liquid free surface. The tangential force is called Marangoni force. Whenever fluid dimensions are less than 10 microns, thermocapillary-driven forces cause the separation of discrete droplets from the main fluid body in terms of propelling them quickly through space.

The occurrence of error originates from two sources. One is due to the self-defects, faults, environmental condition and the operational capability of the system; the other is due to implementation of some specific method or technique. The errors, which occur due to the former reasons, are self-explanatory in the sense that errors in a temperature sensor, for example, occur due to the presence of temperature gradients, sensor nonlinearities, poor thermal contact, calibration drifts, radiant energy and sensor self-heating. In order to obtain better accuracy in a temperature sensor great care in needed. Bear in mind that whenever anyone measures anything, there is invariably some amount of error. This may be due to calibration. Therefore, it is advisable to calibrate the sensed signal periodically. On the other hand, the error that occurred due to implementation of a specific method or technique can be understood through an example as follows. Consider an analog-to-digital conversion process. While converting the analog signal to its digital equivalent a quantization error occurs. Explicitly such type of error is called mean square quantization error (MSQE). MSQE occurs due to quantization process (a method used in Analog to Digital conversion process) and quantitatively it is given by $s^2/12$, where s is the step size.

14.2.5 Accuracy and Precision

Accuracy is defined as the agreement between a measured quantity and the true value of that quantity. In general terms, the *accuracy* of a measurement determines how close the measurement comes to the true value. If the actual value is 1.234 and we say that it is 1.235, then we are precise to the second decimal place but inaccurate by .001. Therefore, it indicates the correctness of the result in terms of implying the degree of agreement of a measured value with the true or expected value of the quantity of concern. Some examples can be given.

- Oscillators exhibit a variety of instabilities due to aging, noise, frequency changes with temperature, acceleration, ionizing radiation and power supply voltage. The term accuracy can be used in describing the quality of oscillations with respect to its instabilities.
- A flowmeter with a range of 20 feet per second and a velocity accuracy of 1% of full scale has an error of ±0.2 fps throughout the operating range.
- Consider a resistor with indicated resistance value of 5.67 ohms at the time of purchasing from a store. Later on it is tested and the measured value is 5.66 ohms. Then the accuracy of the measurement is 5.67–5.66 = 0.01 ohms. Such a calculation gives the absolute deviation of the measurement, which is given by (0.01)/(5.67) = 0.17%.

Accuracy can be determined if some prior knowledge of the true value is available. It is sometimes reflected through the statistical error in the readings that are obtained as a result of the imperfections in the device or instrument. It is also a combination of a number of terms representing uncertainty in the measurement and calibration processes. The influential factors of nonlinearity, hysteresis, repeatability and temperature effects are considered in determining accuracy.

There exist a closely related performance measurable parameter associated with the term accuracy, known as fidelity. The meaning of fidelity is the degree to which a system can accurately reproduce the essential characteristics of the input signal. High fidelity implies high accuracy, however, the primary difference is that the reproduced output only possesses the essential and selective characteristics of the impressed input.

The *precision* of a measurement reflects how exactly the result is determined without reference to what the result means. Although the terms accuracy and precision are frequently used in many places, technically they are not same. These two terms represent distinct performance characteristics. The

difference between the accuracy and the precision is that the former indicates 'how close the number is to the true value' and the latter indicates 'the size of uncertainty'.

14.2.6 Nonlinearity

The systems are modeled based on the input–output relationships. For example, a resistor element (electrical system) can be represented by input–output model equation given by, $i(t) = G \times v(t)$, where $v(t)$ is the input voltage, $i(t)$ is the current and G is a constant. Similarly, the input–output model equation of a thermal system can be written as $RC(d\kappa/dt) + \kappa = \kappa_0$, where $d\kappa/dt$ is the rate of change of output temperature, κ_0 is the input temperature difference between the two points of interest, R is called thermal resistance and C is called the thermal capacitance. The thermal system is a first order system as it has a derivative term. A second order system, which should have a second derivative can be written as $a_2\ddot{y} + a_1\dot{y} + a_0 y = b_0 u(t)$, where the coefficients a_2 a_1, a_0 and b_0 are the constants characterizing the system and $\ddot{y}$ and $\dot{y}$ are the second and first derivatives of the output function, respectively. y and u represent output and input of the system. We distinguish two types of systems, linear and nonlinear. All the above systems are called linear systems since the input–output relationships are linear. Nonlinear systems include nonlinear terms. Followings are the two model equations of some nonlinear systems since they contain the power of the independent variable y.

$$\frac{d^3y}{dt^3} + y^2\frac{d^2y}{dt^2} + y^4 = \frac{du}{dt} \tag{14.1}$$

$$\frac{d^2y}{dt^2} + \frac{dy}{dt} + y^3 + y^2 + y = u(t) \tag{14.2}$$

The basic difference between a linear and nonlinear system is that the superposition principle is applicable only to the former one. That is if, $y_1 = f(u_1)$ and $y_2 = f(u_2)$ then,

$$y = y_1 + y_2 = f(u_1 + u_2) = f(u_1) + f(u_2) \tag{14.3}$$

Many of the real life physical systems are nonlinear in nature. Nonlinear systems are more complex, and much difficult to understand because of their lack of simple solutions. In nonlinear systems the solutions to the equations do not form a vector space.

Consider an example of an amplifier. In certain operational range, the output signal strength of the amplifier may not vary in direct proportion to the input signal strength. This is an example of nonlinear behavior. Quantitatively, in this typical example, the nonlinearity can further be understood as the relative difference between the responsivity at an arbitrary input power and the responsivity at the calibration power. In a system that exhibits nonlinearity, the input–versus–output plot will appear as a curved line over part or all of the input range. Two examples are shown in Fig. 14.1.

The nonlinear function can be expressed by Taylor series expansion, which contains higher order terms since it is a power series. Eliminating the higher order terms from the power series can approximate the nonlinear systems. The degree of closeness depends on the number of terms that are to be eliminated. This process is called linearization. From another perspective, one should note that measurements of the response of a nonlinear system to one or more inputs could be used to obtain an approximate mathematical description of the way in which the response depends upon the input. Such a description can be used to predict the response of the system to inputs other than those used in the measurements.

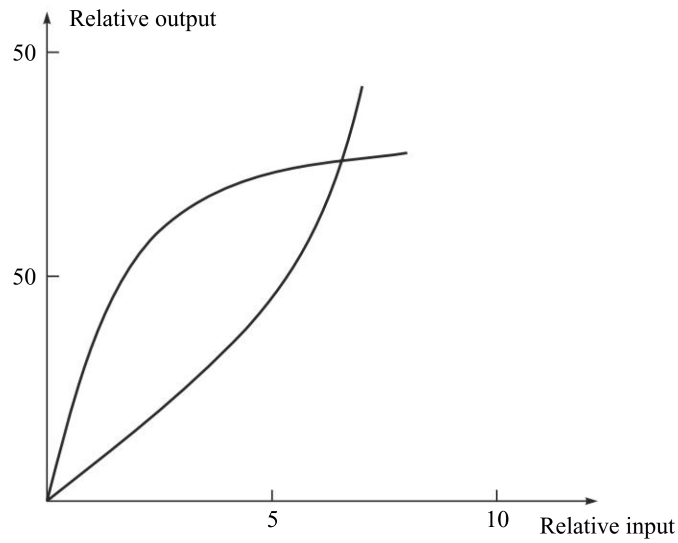


Fig. 14.1 *Typical responses of a nonlinear system*

14.2.7 Reachability

Figure 14.2 shows an example of a block diagram of an open-loop control system. The term 'open-loop' comes from the fact that the output only depends on the inputs. This is a complete system by itself. The control system takes the input from the controller in order to produce output by the action of the plant. The relationship between the input and output are mentioned in terms of transfer function, which is defined as the ratio between the Laplace transform of the output ($Y(s)$) and the Laplace transform of the input ($U(s)$). If the output is proportional to the input, the plant is called a linear system.

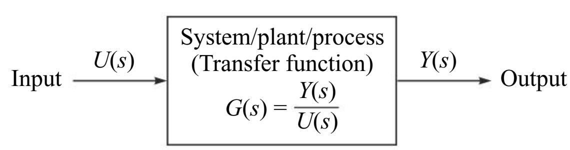


Fig. 14.2 *Block diagram of a system/plant/process*

In a basic open-loop control system, the controller takes the reference input called setpoint and outputs a control signal to the plant or process (Fig. 14.3). This configuration also called feed-forward open-loop control system. The controller is designed and tuned using accurate model of the plant. Any inaccuracy in the system model results discrepancy in the desired output response.

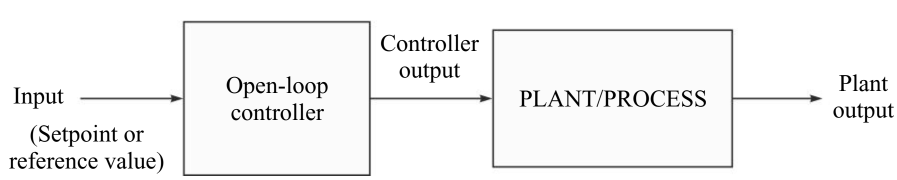


Fig. 14.3 *Block diagram representing open-loop control system*

Figure 14.4 illustrates that if a system with transfer function $G_1(s)$ is connected with another system with transfer function $G_2(s)$ then the overall transfer function of the system is the product of individual transfer functions. In general, in open-loop configuration, mathematically, the overall transfer function of the composite system is given by the following formula.

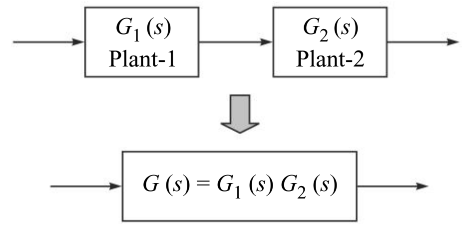


Fig. 14.4 *Equivalence of Open-loop transfer function*

$$G(s) = G_1(s) \cdot G_2(s) \cdot G_3(s)\ldots\ G_n(s) \tag{14.4}$$

A closed-loop control system, on the other hand, uses input as well as some portion of the output to regulate the output. The closed-loop systems are also called feedback control systems. In feedback control the variable required to be controlled is measured. This measurement is compared with a given setpoint. If the *error* results, the controller takes this error and decides what action should be taken to compensate and hence to remove the error. Errors occur when an operator changes the setpoint intentionally or when a process *load* changes the process variable accidentally. The error could be positive or negative.

An automatic speed control system of a typical DC motor is illustrated in Fig. 14.5. In this example, when the output differs from the desired speed, the error signal adjusts the field current of the motor in order to restore the desired speed. Such a closed-loop feedback system typically uses *negative feedback.* If the speed becomes faster than the desired speed a negative value results as the error signal. The negative value causes the DC motor to reduce in speed and therefore compensate for the excess in speed.

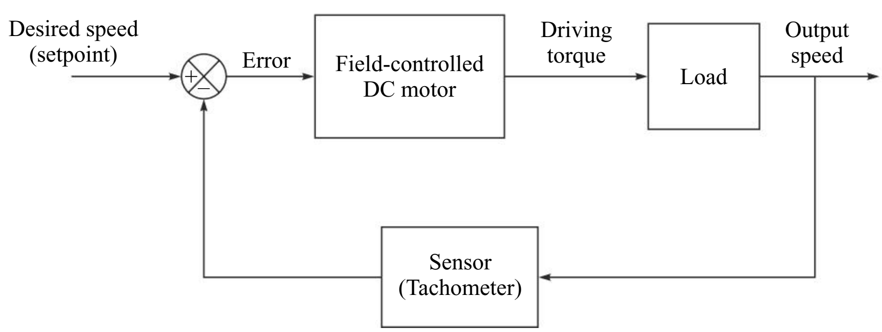


Fig. 14.5 *A typical automatic speed control system of a DC motor*

Because of feedback action, the plant produces stable output. How the stability is maintained through feedback will be explained shortly. Prior to that, it is essential to know the terminology in terms of block diagram of a closed-loop control system.

Certain features are common to all types of closed-loop control systems. These are illustrated in Fig. 14.6. The figure is the block diagram of a typical closed-loop control system. Every closed loop system outputs a controlled variable *y(t)*, which is close to some reference input value *u(t)*. The interference due to external load, however, won't allow the production of the desired value. This is the inherent feature (dynamic properties) of the plant or process. In order to achieve the desired output, the control system has to generate an error signal *e(t)*. This error signal actually regulates the flow of appropriate input (energy) into the controlled system in terms of $c_y(t)$, so as to enable the plant to minimize the error, and thereby compensate the effects of the external load or disturbances.

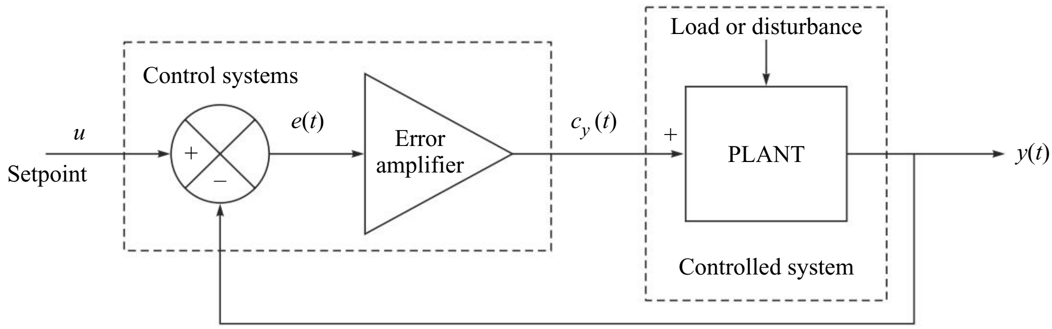


Fig. 14.6 *Block diagram representing closed-loop control system*

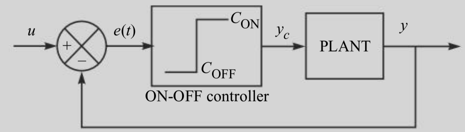
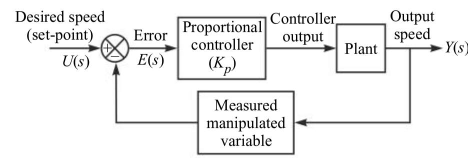
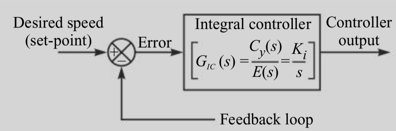
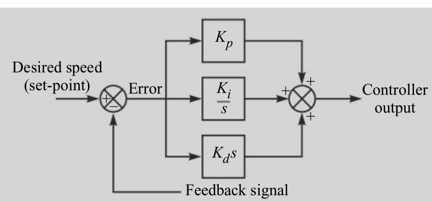
The closed-loop control schematic must have a plant which is to be controlled. The plant is referred to as the *controlled system*. The block that controls the plant (i.e. the controlled system) is called the controller. The controller is not a physical controller as will be pointed out latter. It is a manipulation method that controls and regulates the output through feedback or closed-loop action. The algorithm of the manipulation can be implemented in a *physical controller* in order to achieve the objective, i.e. to regulate the output.

Mostly automatic control systems are based on feedback model. Feedback principle suggests that the output has to be compared to the input to produce the error signal, so that *reachability* can be achieved. Reachability is the desirable or wanted output. Reachability is not guaranteed in open-loop configuration. For which we go for implementing closed-loop or feedback control scheme.

An example can be pondered. To operate an electric motor precisely at an achievable speed of *r* RPM, we have to provide input signal (setpoint) of say, *u* units. Under every circumstance we should expect that the motor should reach at the speed *r* for same value of setpoint without any ambiguities or uncertainty. However, because of the dynamic[2] nature of the system and uncertainty involvement, the reachable speed may not be attended with the same input in every circumstance resulting reachability problem. In order to deal with such problems, the implementation of *automatic control* is necessary. The conventional form of automatic control is the feedback control. The principle of feedback control is based on a kind of on-line reachability test. The feedback system tests the output in real-time, and if it is not the desired value (it could be either positive or negative compared to desired value), then an error signal is generated and is fed to the system to achieve the desired level of reachability. The generation of error signal constitutes as a controller and the act of generation is called control action. In practice the error signal is not directly fed to the plant, rather a refined or modulated version of it is applied. There are various ways of refining the raw error signal. The refining is performed by employing various types of algorithms, which are based on what has been defined as control laws. Some important conventional control laws are proportional (P), proportional plus derivative (PD), proportional plus integral plus derivative (PID) control laws. The control laws are reasonably called controller. This is not a physical controller rather a logical controller. P, PD and PID control actions are governed by their respective fundamental equations as mentioned in Table 14.1. The control equations can be realized by using any programming language and can be implemented in a physical controller such as microprocessor or microcontroller.

[2] Dynamic behavior refers to time-varying property of the system. Here the system is the motor (The system is also called plant).

Table 14.1 *Mathematical representation of various control laws (Controllers)*

Controller	*Equation*	*Graphical Representation*	*Terms*
On-Off	$c_y = c_{ON}$ for $e(t) > 0$ and $c_y = c_{OFF}$ for $e(t) < 0$.		$e_f(t)$ is the error signal, y_{ON} and y_{OFF} are the two control levels for $e(t) > 0$ and $e(t) < 0$
Proportional controller	$K_p = G_c(s) = \dfrac{C_y(s)}{E(s)}$		The coefficient of deviation is called proportional gain, where $G_c(s)$, is the transfer function of the proportional controller, $C_y(s)$ is the Laplace transform of the output of the controller, $c_y(t)$ and $E(s)$ is the Laplace transform of the error signal, $e(t)$.
Integral Controller	$c_y(t) = \dfrac{1}{T_i}\displaystyle\int_0^t e(\tau)d\tau$		where $c_y(t)$ is the output of the controller. $e(t)$ is error signal and T_i is called the integral time.
Derivative Controller	$G_D(s) = \dfrac{C_y(s)}{E(s)} = K_d s$		K_d is the constant of proportionality, usually referred to as derivative time, or simply derivative gain.
Proportional-plus -integral controller	$G_{PI}(s) = \dfrac{C_y(s)}{E(s)} = K_p\left(1 + \dfrac{K_i}{s}\right)$ $= K_p\left(1 + \dfrac{1}{T_i s}\right)$		
Proportional-plus-derivative controller	$G_{PD}(s) = \dfrac{C_y(s)}{E(s)} = K_p(1 + K_d s)$		
Proportional-plus-integral-plus-derivative controller	$G_{PID}(s) = \dfrac{C_y(s)}{E(s)}$ $= K_p + \dfrac{K_i}{s} + K_d s$		where K_p, K_i and K_d are called proportional, integral and derivative gains of the controller, respectively.

Thus if a feedback loop is added to the control system, the controller gains the ability to react to the quality of the response of the system output. In this way the system can have better chance of performing as desired. This is the reason why the feedback control systems are more stable than open-loop control system. Note that the feedback concept is fundamental to all sorts of conventional automatic control system.

14.3 TRANSIENT PROPERTY

If at time $t = 0$, an input forcing function is applied to a system we expect that the output should respond immediately at the same time. However, in reality, the system takes time to produce the output response. This delay or lagging characteristics is due to the presence of energy storing elements within the systems. The output starts from initial value (the initial value could be zero) and reaches the *steady-state* value after certain time. How the output reaches the steady-state value solely depends on the type of the system in hand (e.g. first order, second order, linear, nonlinear, etc.) and the nature of applied forcing function. This is called transient property of the system. The property is specified through what is known as *transient-response* specification (TRS). The transient-response specifications involve the following terms.

- Rise-time
- Peak-time
- Settling-time
- Steady-state value
- Maximum overshoot

The above terminology is based on a step input to the system. Step function is an ideal forcing function to test a system. Figure 14.7 illustrates the transient response corresponding to a step function, as the applied forcing function.

Rise-time is the time required for the output response to reach 90% from 10% of the input value. *Peak-time* is the time required to reach the peak overshoot value. *Settling-time* is the time at which the response attends within 98% of the final value. An error limiting 2% is called tolerance. The output response, settles within this tolerance band only after the settling time (t_s) is reached and the level of output is referred to as initial steady-state value. Initial steady-state value starts at $t = t_s$. The final steady state value refers to the response value at time $t = \infty$. The maximum overshoot is the difference between the maximum peak and the input value. It is usually expressed in percentage. Figures 14.2(b) and (c) illustrate the TRS of two other systems or the same system with the addition of feedback control.

14.4 HYSTERESIS

If someone coerces something it *yields* and when s/he *releases* the driving force, it does not give back the way it was yielding then the yielding and releasing responses will follow different path. The response curve will be similar to that shown in Fig. 14.8. This curve is known as hysteresis curve. Hysteresis is thus a dual-output phenomenon for the same input.

The curve is actually an illustration of a magnetic hysteresis. Magnetic hysteresis occurs when a material with high permeability such as soft iron is magnetized. The magnetization is performed by an external magnetic field ***H***, called magnetizing force. If a magnetic field is applied to a previously unmagnetized soft iron sample continuously and subsequently the field is removed, the soft iron sample

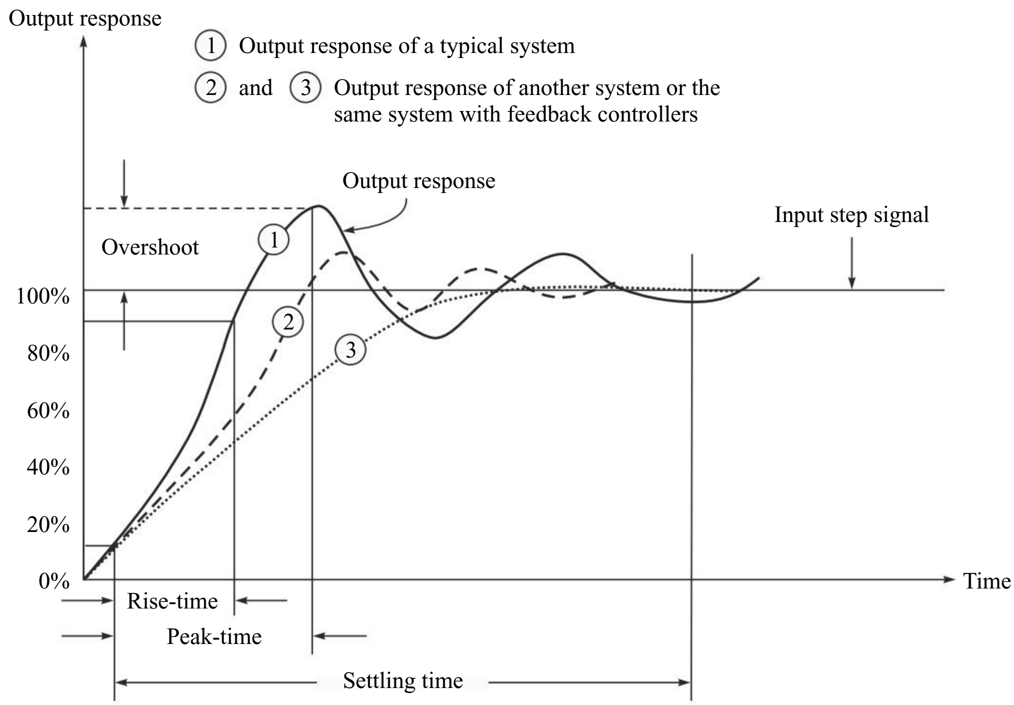


Fig. 14.7 *Transient-response specifications*

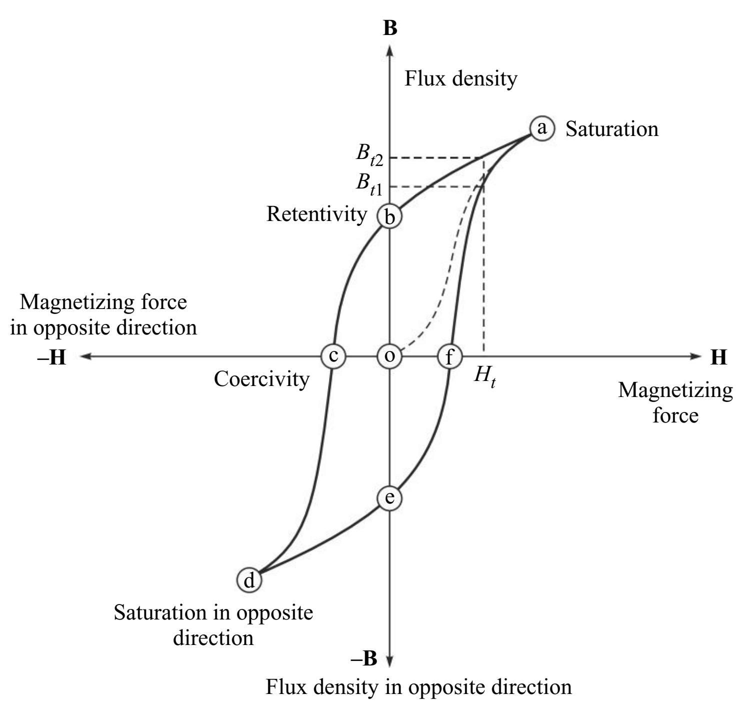


Fig. 14.8 *Magnetic hysteresis*

retains a residual magnetization and it becomes a permanent magnet! The graph of the magnetic induction ***B*** within the sample versus the applied magnetic field ***H*** will follow the dotted path shown in the figure. When the magnetizing force ***H*** is zero, the magnetic induction is also zero. As ***H*** increases, ***B*** follows the dotted path and reaches a point called *saturation* point, a point beyond which no further magnetization is possible even if the ***H*** increases to infinity. Now if we decrease ***H***, then ***B*** will also decrease but it will follow another path intersecting B-axis at *retentivity*, a point at which the magnetization still persists even though the value of ***H*** is zero. As ***H*** further decreases in opposite direction the curve will pass through *coercivity* point (along the –ve ***H*** axis) and then will reach at the negative saturation level. If ***H*** is increased the curve will follow a path touching the points e, f and finally arrive at the point 'a' again. If the cycle is repeated the resulting plot will still follow a path *abcdefabc*... This curve is called hysteresis loop.

Mathematically, the hysteresis response due to the external influence is eventually a doubled-valued function. One value applies when the influence is increasing, the other applies when the influence is decreasing. For example, for same H_t the magnetic induction values are B_{t1} (yielding condition) and B_{t2} (releasing condition). Theoretically, this physical phenomenon depicts that the yield depends not only on the present magnitude of that influence but also on the previous history of the system. Thus, hysteresis represents the history dependence of physical systems.

Some sensors/transducers possess hysteresis effect. The sensor provides one reading when the external influence is increased and shows another reading when the influence is decreased. There now exist a difference in measurement between the sensor output readings for the same applied load. The error caused by this effect is called hysteresis error. Hysteresis error degrades the sensor performance.

In another context, however, the effect of hysteresis is barely necessary. For example, consider a thermostat-based heating system. Such a system uses ON-OFF type controller. When the output is lower than the setpoint the controller is turned ON (i.e. provides an ON output), and once the output is more than the setpoint the controller provides OFF output. The turn-ON and turn-OFF points are deliberately made to differ by a small amount called *dead-band* (Fig. 14.9) to prevent noise from switching the controller unnecessarily when the output is nearly at the setpoint. This design in hysteresis prevents the output from switching from OFF to ON too rapidly. The hysteresis is designed into the control action between the points at which the control output switches from OFF to ON. If the hysteresis is set too narrow, rapid switching will occur. Therefore, the dead band and hence the hysteresis should be set so that there is sufficient time delay between the ON and OFF modes of the outputs. The sensitivity of the On-Off controller depends on the dead band (hysteresis).

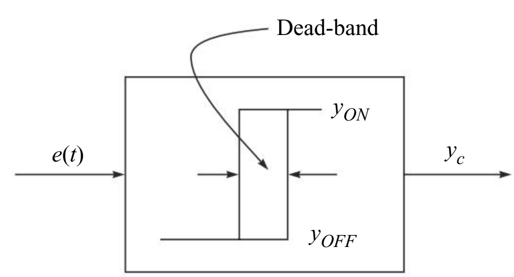


Fig. 14.9 *On-off controller with hysteresis or dead-band: e(t) is the input signal to the controller. Y_c is the controller output (either y_{ON} or y_{OFF})*

14.5 STABILITY

The stability of a system is often extremely important and sometimes is a safety issue in the engineering field. The very accurate and theoretical definition of stability is well known to control engineers. The

general meaning of the term 'stable' however, motivates someone to understand it as resistance to change. Stability can be defined as the ability of a component, circuit or system to maintain a fixed level of operation within specified tolerances under varying external conditions and disturbances, such as voltage, frequency, temperature, longevity, and so on. The stability in fact relates to its response to inputs and external disturbances.

A system, which remains in a constant state, can be considered to be stable. This equally implies that if the system is disturbed by an external action and subsequently returns to the constant state when the external action is removed, then the system is said to be stable. An illustration of stability can be best understood from the graphical presentation shown in Fig. 14.10. The first plot is the input to the system. In this case the input is a step signal. Other plots are the outputs in response to the input. Depending the nature of the system the output may vary with respect to the same input.

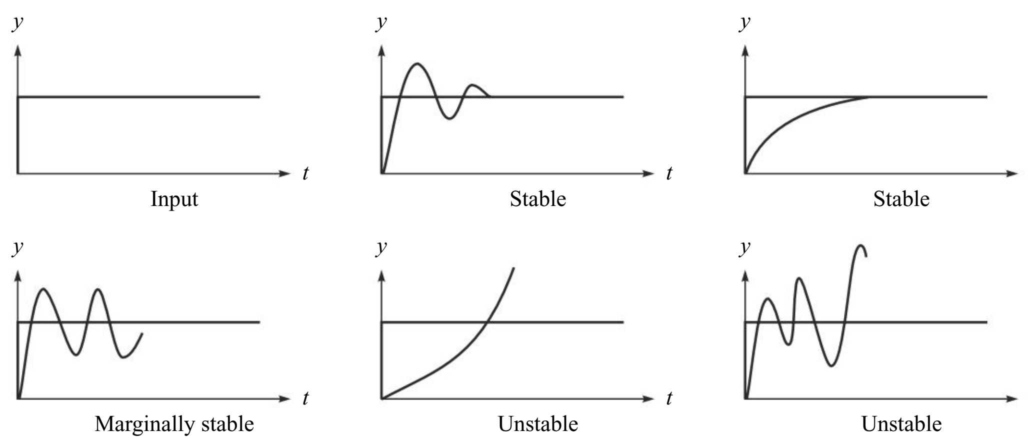


Fig. 14.10 *Plots showing types of stability responses resulting from an input*

14.6 RELIABILITY

Reliability provides the theoretical and practical means whereby the capability of devices performing their required functions for desired periods of time without failure can be expressed. The critical applications, with which many products and systems are entrusted, consider reliability as an essential attribute. Reliability is the study of prediction, estimation or optimization of the life of the products or systems. Issues of safety, product liability and warranties are all strongly dependent on reliability. Reliability is mostly a measure of consistency and it concerns with probability. The probability that a device, product or system would perform a specified function under specified operational and environmental conditions during a specified time is quantified to account for reliability. The specified times are of three types given by

- An instant of time
- A time interval
- All the time

14.6.1 Reliability Curve in General

Failure rates of most products show typical characteristics that are plotted in Fig. 14.11. The curve takes product life on the x-axis and failure rate (FR) on the y-axis. Life can be expressed in minutes,

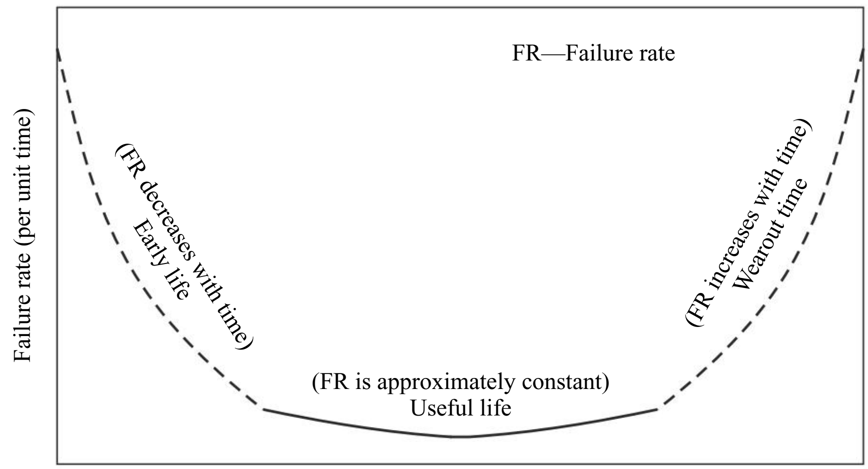


Fig. 14.11 *A reliability curve*

hours, years, cycles, actuations or any appropriately defined quantifiable unit. The FR is specified as failures per such unit (usually time). The curve has three regions: early, useful and wearout region. Notice that most of the products begin their life with a large failure rate. The reason is due to three factors such as:

- Manufacturing defects
- Poor workmanship
- Poor quality control

During useful life the FR is nearly constant. The FR increases during wearout period. The duration and hence the proportionality between these three regions vary from system to system, however the nature of characteristic curve remains similar.

14.6.2 Reliability of MEMS

The poor reliability of the MEMS devices evidently comes from the manufacturing defects. MEMS manufacturing defects are closely related to a term called *fatigue* which is a failure mode associated with crack initiation and growth driven by time-varying stresses. The meaning of fatigue is simple to understand, for example, when a motion is repeated the component that is doing the work can become weak. Time-varying stress drives the MEMS components towards failure region (Fig. 14.12), initiating crack and its growth. The rate at which a crack grows has substantial impact on the life of a component. The propagation of a crack occurs during the subsequent step of fatigue failure. It is obvious that as a crack begins to propagate, the size of the crack also begins to grow. The rate at which the crack continues to grow depends on the unwanted stress level. The rate at which a crack grows can be expressed mathematically as follows.

$$\frac{dL}{dc} = \mathcal{A}(\Delta K)^p \tag{14.5}$$

where the variables $\mathcal{A}$ and p depend on the properties of the material. dL is the change in crack length, and dc is the change in the number of cycles.

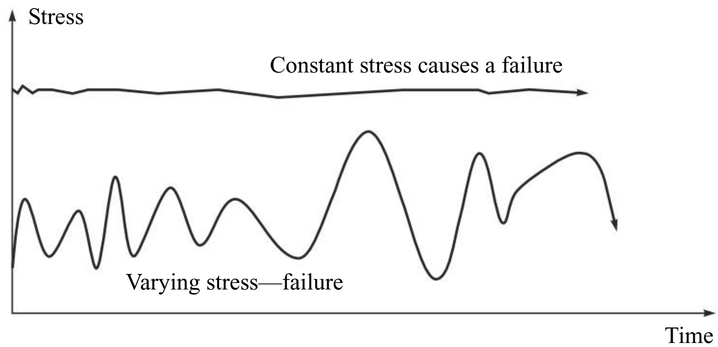


Fig. 14.12 *Alternating stresses cause failure at reduced stresses (Source: MEMS Reliability Newsletter, 1/1, 2001)*

It should be pointed out that material strength is not a key limiting factor in MEMS reliability. The essential requirement is that the material should sustain high stress so that it can be used in joints, beams and springs.

Reliability of MEMS has become the overriding priority for industry and focused on delivering low-cost products to the market in the shortest possible time. Reliability department is very commonly seen in most manufacturing sectors. The key reason for opening reliability-engineering wing within the MEMS design activity is that the products are becoming more complex with the addition of more components and features. The advantages that can be gained by implementing reliability program are summarized as follows.

- Optimum warranty period
- Preventive replacement time
- Study of the effects of age
- Estimation of the required redundancy
- Increase of customer satisfaction
- Promotion of device and company reputation
- Productivity

14.6.3 Remedy

While addressing the reliability issues, attention has to be given to the following points. MEMS are the combination of circuits and micro-machineries. The reliability aspect matters with both the electronic and the mechanical parts, complicated by their interactions. Selective materials and their composites must be chosen for appropriate components. Iterative test procedure must be conducted in order to validate the compatibility of the material to be used for development of such component. Broadly, the following proof-of-design concept must be taken into account while dealing with the reliability issues.

- Material property characterization
- Process development and characterization
- Design methodologies
- Design for manufacturability (DFM)

An example can be given in this respect. In a capacitive RF MEMS switch, the most important failure mode is parasitic charging of the dielectric material (corresponds to material property characterization). An improved analytical model that can enable the calculation and understanding of the effects of insulator charging on the behavior of capacitive RF MEMS switches to describe the way they fail should be addressed (corresponds to process development and characterization). Emphasis should also be given on studying the effect on various design structures with a wide range of pullout voltages

(corresponds to design methodologies). Last but not the least DFM must be considered as a blueprint as it provides valuable information based on previous knowledge. Table 14.2 illustrates some commonly encountered failure modes, causes and the failure analysis techniques in a broader sense.

Of these, the characterization of material plays an important role. The design considers automated tools to assist in studying the behavior of the MEMS structure in order to improve reliability.

Table 14.2 *Failure modes, causes and analysis techniques*

Failure modes	*Causes*	*Failure analysis techniques*
1. Extrinsic contamination	1. Mishandling	1. Scanning Laser Microscopy (SLM)
2. Component fusion	2. Capillary forces	2. Scanning Electron
3. Sticking	3. Operational methods	Microscopy (SEM)
4. Clamping of movable	4. Mechanical instabilities	3. Focused Ion Beam (FIB)
parts like gears,	5. Electrical instabilities	4. Atomic force microscopy (AFM)
links, joints	6. Interferences	5. Laser Cutting
5. Static overload		
6. Creep		
7. Environment		

14.6.4 Material Characterization

Characterization is a means of advocating the merits and demerits of selected materials as well as the selection of the structural configuration for a particular MEMS component. MEMS and NEMS necessitate sophisticated material characterization (MC) techniques. Characterization of material properties, microstructures, and residual stresses is critical to the understanding of mechanical behavior, design performance of MEMS and NEMS. MC tool characterizes structure, structural defects and properties for research as well as micromanufacturing. The procedural methodology is an integral part of materials research, development, application, and reliability assurance. The characterization especially emphasizes the need to estimate residual and critical stresses, material physical properties, material damage evolution, and degradation assessment, material processes and microfabrication, and defects at any location. The important points that are considered in order to achieve the reliable products are,

- Selection of appropriate materials. For example in some application ceramics may be preferred instead of metallic or polymeric kind. Further, although silicon and nitride are more fatigue resistant, these materials are less resistant to brittle fracture in the presence of existing defects.
- Since the materials existing in the environment are susceptible, consideration must be given to the operating environment. For instance, water affects ceramics, metals, and polymers. Further, organics also affect polymers.
- Calculate the desired ultimate stress. Design the component for stresses well below, at least 3–4 times below, the calculated ultimate strength.
- Develop smoother surfaces (flat, curved, etc.) and sidewalls by using appropriate processes.
- Interconnections, interfaces, links, anchors and joints must be simulated and analyzed beforehand. These parts of the designs bear loads and stress, and therefore could promote crack initiation.

- Validation testing must be carried out. This testing is essential due to the reason that, the evidence of materials property from data sheet may not match with the practical design. For example, the basic mechanical properties of thin films such as Young's modulus, Poisson's ratio and tensile strength are not completely characterized. The lack of material data requires thorough testing of each component. Fatigue specimens must be prepared and tested in order to observe the failure in the section of interest under well-defined loading conditions.

In MEMS devices inertia factor is of little concern, however, the effects of atomic forces and surface science dominate the reliability issues.

14.7 AVAILABILITY

Availability and reliability are two sides of a coin. More reliable implies more available. However, qualitatively availability is defined as the ability of the device to perform its implied function at a stated instant or over a stated period of time. Availability can be considered as the quantitative expression of reliability. The quantitative parameters are, Mean Time Between Failure (MTBF), Failures In Time (FIT) and Mean Time To Repair (MTTR).

All the above quantitative estimates are based on the degree to which the system will be operative. MTBF is the average time between failures of the device. MTBF information can be obtained from manufacture's data sheets. FIT is the total number of failures in a billion hours. It is a more intuitive way of representing MTBF. The MTTR is the time taken to repair a failed device. It is the probability of the device not being in the failed condition at some randomly chosen moment in the distant future. Usually, the MTTR in case of microsystems are negligible due to the fact that the components within the MEMS devices are considered to be irreparable. Mathematically, the MTBF can be expressed as:

$$\text{MTBF} = 1/(FR_1 + FR_2 + FR_3 + \ldots FR_n) \tag{14.6}$$

where FR is the failure rate of each component of the system containing n components. Once *MTBF* and MTTR are known, the availability can be expressed quantitatively as,

$$A = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} \tag{14.7}$$

14.8 TRACEABLE CALIBRATION

Calibration is the process of comparing the measured signal with reality. The accurate measurement of measurand and hence the calibration is vital in the sensing fields. Consider a servo controlled system, where sensors are used as feedback elements. Until recently only the continuous contact between a measurement and the output controlled system has been the unidirectional flow of measurement data. It has been presumed that the sensor elements are viewed as simple signal generators and their data are assumed to be correct. This single stream of information is used in a variety of ways: for monitoring the process, for feedback control and also for ensuring safety through the use of hard-wired trips.

As the effective operation of the measurement system depends completely upon the data received from the sensors any failure leading to inaccurate measurement is undesirable and sometimes intolerable. Some systems therefore require sensory devices to be subjected to periodical recalibration. An important aspect of applying dynamic calibration to any sensory devices is referred to as *traceable calibration*. Traceable calibration is a concept that ensures that all measurements made are accurate and

valid. Sensors employing traceable calibration algorithm are called *smart sensors*. The concept is also defined as *sensor validation*. Sensor validation is taken to imply the generation of validity index of the sensed signal. Along with measured value a truthfulness value or more accurately the *validity index* (VI) is generated as shown in Fig. 14.13. The validity index simply represents the quality of the sensed signal. Note that the validity index is a function (inversely proportional) of unwanted noise or faults. The quality can be expressed in terms of percentage. If the validity of the signal is 90% the control system can accept the sensor output, else not. Such types of sensors are useful for hard, real-time critical applications. The validity index can be generated by employing methods such as model-based, spectrum analysis and sensor fusion technique.

- *Model-based:* Model based approach is very useful in those systems where many sensors are employed for various measurements. This approach is also called analytical redundancy method. The quantitative analysis of data provides a concise but informative definition of the underlying happenings in the form of a mathematical model. The procedure involved in this scheme is as follows.
    - Obtain the signals from several sensors in a system,
    - Develop a model including process dynamics of the sensory system,
    - Look for inconsistencies
    - Provide validation index
- *Spectrum analysis:* Spectrum analysis scheme utilizes the frequency domain analysis. Typical sensed signals have a nominal frequency band. Under normal operating conditions the sensor output produces typical frequency spectrum. Any abnormality on the spectrum indicates faults and anomaly.
- *Fusion:* When identical sensory devices at least triplex in configurations are verified by comparing the measured value at a given instant and the failures are isolated by majority voting, then it is called sensor fusion. Fusion techniques are called hardware redundancy scheme since faulty sensor data are eliminated and isolated by majority voting. In order to achieve sensor fusion schemes, multiple sensors have to be fused together; called arrayed sensor, and the output is taken from the non-faulty sensor(s). In MEMS, sensor array is preferred because of its simple implementation and integration methods.

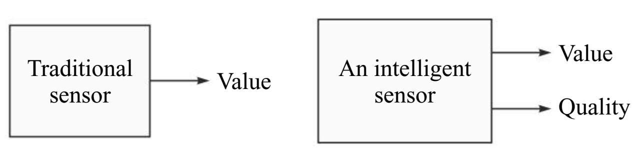


Fig. 14.13 *Typical traditional and intelligent sensors*

14.9 TRIMMING

Trimming is a process of removal of unusual portions prior to its final production. This cutting down of various components or element of the MEMS devices is carried out during fabrication process. Sometimes trimming is referred to as tuning, although the very usefulness of the term tuning is found in RF MEMS application. One way of achieving the job is through laser technology, hence the method is called laser trimming. The micromachining system is always integrated with the trimming system. In particular, trimming allows adjustment and optimization of parameters, such as sensitivity, zero pressure

offset and temperature response. As an example, a calibration circuit within the MEMS device can be perfectly set-up by trimming the resistance elements constituting the circuit. Such trimming process is defined as resistance trimming technique. Trimming improves the performance. Trimming is relatively an old process and rarely used in the present technology.

14.10 IMPERFECTION COMPENSATION: AN EXAMPLE

Many of the MEMS devices have structural layers, which may be made up of composite metal-dielectric materials. These layers experience larger vertical stress gradient compared to their counterpart such as homogenous polysilicon. When subjected to actuation the metal-dielectric materials based layers tend to curl out-of-plane, which is undesired. The curling should be in-plane during operation. A badly designed MEMS device having suspended proof-mass fingers and stator fingers may curl in opposing directions, as shown in Fig. 14.14(a). If the fingers (layers) have been designed for capacitive sensing applications, the desired sensitivity may not be achieved since sidewall-sensing capacitance is significantly reduced due to out-of-place curling effect. Further, the sidewall capacitance may change with temperature. The curling appears due to structural imperfection. In order to achieve the in-plane curling in such applications, appropriate structural design principles must be adopted. The structural imperfection has to be compensated through additional design considerations. The improved design technique as shown in Fig. 14.14(b), can take the advantage of compensating the out-of-plane curling effect. In this typical design, the stator (right-side component) and proof-mass fingers (comb fingers) are anchored along a common axis. The stator is usually connected to a cantilevered frame, which is very rigid as compared to the proof-mass suspension. The fingers now can curl in-plane eventually providing maximum sidewall capacitance. By employing appropriate structural design principle in terms of implementing imperfection compensation technique, performance of a system can be significantly improved.

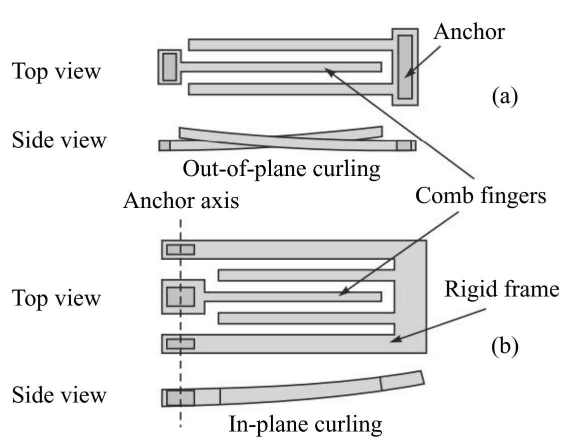


Fig. 14.14 *An example of imperfection compensation technique (Curl compensation for comb-fingers) (a) Traditional design (b) Structural compensation for curl matching. (Source:* Zhang et al., The Robotics Institute Carnegie Mellon U. and Motorola Inc.*)*

14.11 ARRAY DEVICES

The purpose of fabricating array configured MEMS devices is multifold. Besides the usefulness of array configuration for achieving a generic validated signal at the output of a sensor, the interest in array also lies with other prospects and viewpoints as described below, with examples.

14.11.1 Read/write Head Array

MEMS-based data storage design consists of an array of extremely small read/write heads. These are called probe tips and are held under a rectangular sled coated with magnetic recording media. Data are stored and retrieved by positioning and moving the sled over the probe tip array while the tips transfer data. Data access is accomplished by moving the media at a constant velocity in the Y direction while data is read or written by the stationary probe tips. MEMS storage chips can be built using standard photolithographic CMOS processes.

14.11.2 Free-space Optical Switch Array

Fiber optic network technology will benefit from using MEMS optical switches. High-Speed Internet (HIS) connections are demanding integrated All-Optical (AO) components that make it possible to attain data transmission rates in the range of tens to hundreds of Gigabits[3] per second. A single optical fiber is capable of carrying 300 Gbps, which is equivalent to approximately 5 million telephone conversations at once. The AO system depends on the optical switching technology requiring high-volume and high-speed data transmission. The light waves carrying data transmitted in the optic fibers must be switched to other fibers as fast as possible.

Free-space optical switch includes an MEMS mirror array for switching the optical transmission path. The switching is performed by tilting a mirror so that the incident optical signal can be directed to a desired path along which another optical fiber is lined up (Fig. 14.15). An array of mirrors can be fabricated in one wafer in order to switch many optical signals simultaneously.

The device has two mirror arrays, namely the input mirror array and the output mirror array. Both the arrays meet head-on an input fiber port array and an output fiber port array integrated with collimating bass lenses, respectively. The input array and the output array are integral with each other and divided from each other by at least two intersecting lines. In the current high-speed optical network scenario, arrays as large as 1000×1000 micromirrors will be needed for switching purpose.

The important considerations while designing the MEMS mirror arrays for switching applications is the way of actuating the mirrors for precise positioning or tilting. In practice, when an electrostatic voltage is applied, the resulting electrostatic forces pull the mirror thereby letting it tilt. The precision of mirror tilting has been a more stringent design. Estimated precision calculation shows that a distance of 500 microns and a fiber core diameter of 2 microns require 0.1° precision as far as positioning of micromirrors are concerned. As arrays become larger, free-space optical path distances increase and very small error in mirror tilt will result in large errors in beam path causing the signal to miss the targeted fiber port. A two degree-of-freedom (2-DOF) with 3D arrangement of micromirrors which can be precision controllable to varying angular deflections with feedback control is desirable for larger optical switching arrays. A schematic configuration is shown in Fig. 14.15. The mirrors work in unison with 2 arrays of optic fibers with collimating ball lens.

[3] Convention: Gb = Gigabits; GB = Gigabytes; Mb = Megabits; MB = Megabytes; 1 byte = 8 bits; Gbps = Gigabits per second; GBPS = Gigabytes per second.

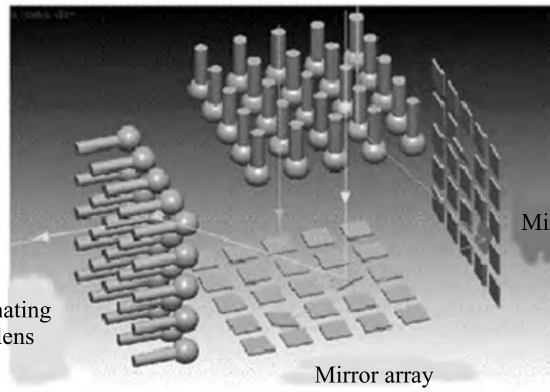


Fig. 14.15 *A configuration of bi-directional 3-D optical switching method (Courtesy:* UC Irvine Microsystems Lab*).*

14.11.3 Grating Light Valve Technology

In order to display an image it is first converted to electrical signal. The equivalent electrical signal is then used to modulate a light source for producing the image on the screen. Light modulators can achieve this. Silicon Light Machines has developed a unique display method for the projection of imago-electric signal using its Grating Light Valve (GLV) principle based MEMS technology. The GLV principle utilizes diffraction phenomenon in order to produce gray-scales of light. The technology uses one-dimensional arrays, i.e. single columns of aluminium-coated silicon nitride ribbon-like strips fabricated with air gap (comparable to the wavelength of the light) over a tungsten-oxide layer of a silicon wafer. Each ribbon is about 100 μm long, 100 nm thick and typically about 3 μm wide, placed alternatively in order to allow a gap between them. The gap between them makes it possible for the structure to behave as a diffractive grating. A pixel is defined in terms of a group of consecutive ribbons. Sometimes, four consecutive ribbons are chosen to make up a pixel. Each ribbon can be pulled down a controlled distance into the air gap by electrostatic actuation. At resting state the aluminium coatings act as mirrors, reflecting the light. If alternate ribbons in a pixel are pulled down, a square-well diffraction grating is formed. The deflection towards the air gap is in the order of the wavelength of light. Light waves falling on the ribbons and reflecting off the adjacent up and down ribbons, due to electrostatic actuation, are out of phase with each other. The phase differences depend on the wavelengths of the light components. The air-gap integrated ribbon-mirror construction is such that the incident light waves interact with the grating structure in a way that causes each frequency of light to radiate from the pixel at a different angle. So the falling light can be modulated by essentially varying the following parameters.

- Width of separation
- Width of the ribbon
- The degree of pull-down

The modulation of light refers to the angle at which certain frequencies should be reflected. This in turn implies the generation of a single color of light out of white light (many colors).

14.11.4 Digital Light Processing Technology

DLP™ technology is also a powerful and flexible display technology. The heart of these display solutions is the Digital Micromirror Device™ (DMD™), a semiconductor-based *light switch array* of thousands of individually addressable and tiltable mirror pixels. DMD, a micromechanical silicon chip, is considered as a spatial light modulator (SLM). It contains millions of tiny, movable aluminium mirrors and electronic logic, memory and control circuitry. Each 16-μm^2 mirror DMD consists of three physical layers and in between them two air-gap layers are present. The air-gaps separate the physical layers so that the mirror can be tilted +10 or –10 degrees, when subjected to electrostatic voltage. When voltage is applied to the addressable electrodes, the mirrors tilt ±10 degrees signifying ON or OFF in a binary signal fashion. There is a projector (light source) and display (screen) unit within the system. The micromirrors are mounted in such a way as to ensure that the tilt is either towards the light source in a projection system (ON) or away from it (OFF); creating a light or dark pixel on the screen surface, respectively. The bit-streamed image code (electrical equivalent of image signal) to be displayed first enters the DMD which is then directs each mirror to switch ON and OFF up to several thousand times per second. When a mirror is switched ON more frequently than OFF, it reflects a light gray pixel and a mirror that is switched OFF more frequently reflects a darker gray pixel. In this way, the mirrors in a DLP projection system can reflect pixels in up to 1,024 shades of gray to convert the video or graphic signal entering the DMD device into a highly detailed gray-scale image.

14.11.5 Calorimetric MEMS Sensor Array

All physical, chemical and biological reactions occur with change in heat. Sensing mechanism may be provided to access the thermal characteristics of such reactions. By this way reaction monitoring is possible. The sensing mechanism provides feedback information on the reaction. Depending upon the nature of the feedback information the reaction can be controlled, e.g. it can be made ON or OFF. Typically, arrays of feedback sensory elements are engaged to obtain the thermal characteristics from the entire volume of the reactants. This in turn allows for spatial patterning or imaging of various types of chemical and biochemical reactions.

The sensing and controlling of reactions is possible with the use of micro-fabricated device array. Oak Ridge National Laboratory has developed the micromachined sensor arrays by CMOS processes, in which as heat is released or absorbed, the device temperature will change with a corresponding change in voltage. The reactions that occur on or near the surface of the device are monitored in real-time with picojoule precision. Conversely, by controlling the temperature, specific reactions can be enabled or disabled. Arraying of these structures would allow control of a variety of reactions permitting the realization of detailed spatial patterns of micro heating.

Initially, the arrayed device was intended for sensing and controlling of chemical and biological reactions those produced by glucose and cholesterol. The biological macromolecules, such as proteins or nucleic acids can directly be attached to these arrayed structures and their reactivity can directly be monitored effectively. The device can also analyse nonlinear oscillating chemical reactions in real-time. Numerous other applications of such an integrated thermometric device array can be envisioned.

14.11.6 Frequency-selective Surfaces

Frequency selection is defined as selection of desired frequency out of many components. Frequency can be selected by using Frequency Selective Surface (FSS). FSS is usually a conducting sheet, understood as arrays of metallic patches which are periodically perforated with apertures. FSS are thus

filters and are primarily used to filter the electromagnetic (EM) waves. The FSS rather function as spatial filter. Figure 14.16 shows an FSS with incident, reflected and transmitted EM waves. Figure 14.16(b) shows another FSS that can select a particular frequency. The EM waves can incident on the surface of FSS at an angle. FSS have been widely used in antenna system of radar and broadband communications.

At the microscopic level, FSSs are composed of ferromagnetic dipole element arrays (FMDEA). FMDEA is fabricated based on electrodeposition of highly magnetic materials by incorporating MEMS techniques. The ferromagnetic arrays behave as magnetic microactuators. The microactuators are actuated by an off-chip magnetic field source. By rotating the dipole elements from 0° to more than 45°, the filtering response of the frequency selective surface can be tuned over GHz range. The device can dynamically attenuate the frequency bands through appropriate configuration and positioning of FSS elements. Certain frequencies of the incident wave are allowed to pass through the screen and others are reflected by properly activating the microactuators.

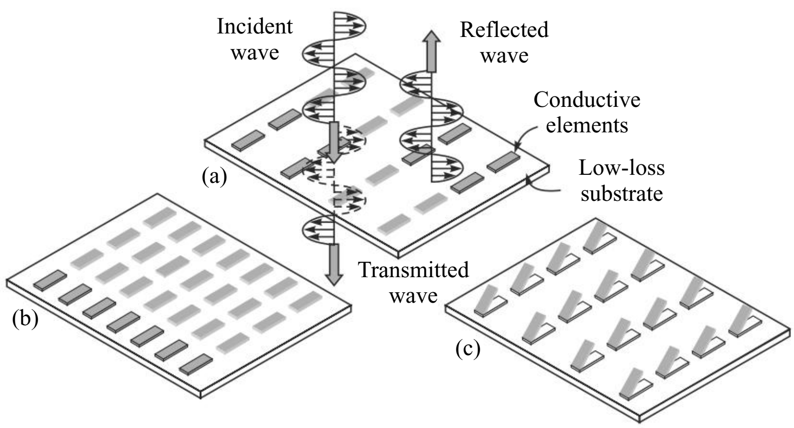


Fig. 14.16 *Frequency selective surfaces: (a) The incident, reflected, and transmitted electromagnetic waves through the FSS; (b) A typical FSS with a different frequency response; (c) A typical FSS with rotating dipoles*

14.12 SUMMARY

Performance is defined as the operational and support characteristics of a system that allow it to effectively and efficiently perform its assigned job over a given period of time. This in turn implies that it is a measure of how well a system will operate. Performance of an engineering system depends on many factors including design parameters, environment where the system is operation, the material from which the system has been designed, the manner in which the system is being operated (duration, manually/automated), etc. Since many aspects are involved in order to mimic the performance, the engineers must be fully aware of the parameters that imitate the performance. Those parameters are called performance measuring parameters or indices.

In this respect, exclusive study materials are included to grasp the knowledge on these indices vis-à-vis performance of a system. In essence the performance measuring parameters are repeatability, resolution, sensitivity, accuracy and precision and nonlinearity. All these parameters have been defined and explained. The knowledge on transient properties of dynamic system is essential. As a result the

transient response specification is introduced. Hysteresis also determines the performance factor. The fundamentals of hysteresis are explained through an example. Reliability and availability of the system can be improved based on statistical reports. In this respect study of the nature of reliability curve is important. The advantages that can be gained by implementing the reliability program within the manufacturing sectors are outlined. Device array plays a major role as far as improvement of performance is concerned.

Points to Remember

- The quality characterizes a particular aspect, capability or attribute of a system, which are separately quantified by numerical values through respective definitive terms. These terms are called performance indices.
- Performance is perhaps more difficult to define but is frequently understood as a particular set of expected effects that may fulfil the envisaged outcome.
- The commonly encountered performance measuring indices are repeatability, resolution, sensitivity, accuracy, precision and nonlinearity.
- Repeatability is the ability to perform the given function iteratively.
- Resolution describes the degree to which a change can be observed and detected.
- Sensitivity is defined as the responsiveness of the variable of interest to an independent forcing variable. Quantitatively it is defined as the ratio full-scale output to the rated capacity of a given transducing or an actuating system.
- The occurrence of error originates from two sources. One is due to self-defects, faults, environmental condition and the operational capability of the system; the other is due to implementation of some specific method or technique.
- Accuracy is defined as the agreement between a measured quantity and the true value of that quantity. In general terms, the *accuracy* of a measurement determines how close the measurement comes to the true value.
- Accuracy can be determined if some prior knowledge of the true value is available. It is sometimes reflected through the statistical error in the readings that are obtained as a result of the imperfections in the device or instrument.
- The *precision* of a measurement reflects how exactly the result is determined without reference to what the result means.
- The difference between accuracy and precision is that the former indicates 'how close the number is to the true value' and the latter indicates 'the size of uncertainty'.
- Nonlinear systems include nonlinear terms. The nonlinear function can be expressed by Taylor series expansion which contains higher order terms.
- Feedback principle suggests that the output has to be compared to the input to produce the error signal, so that *reachability* can be achieved.
- The transient-response specifications defines Rise-time, Peak-time, Settling-time, Steady-state value and Maximum overshoot.
- Hysteresis is a typical dual-output phenomenon for the same input. Mathematically, the hysteresis response due to the external influence is a doubled-valued function.
- Magnetic hysteresis occurs when a material with high permeability such as soft iron is magnetized.
- In some situation hysteresis is unwanted. However, the effect of hysteresis is necessary in some control actions.
- The stability of a system is often extremely important and sometimes is a safety issue in the engineering field.

- A system, which remains in a constant state can be considered to be stable.
- Reliability provides the theoretical and practical means whereby the capability of devices perform their required functions for desired periods of time without failure.
- Failure rates of most products show typical characteristic curve. The curve has three regions: early, useful and wearout region. It is observed that most of the products begin their life with a large failure rate.
- The key reason for opening reliability-engineering wing within the MEMS design activity is that the products are becoming more complex with the addition of more components and features.
- Characterization of material properties, microstructures, and residual stresses is critical for understanding mechanical behavior and design performance of not only MEMS, but also for NEMS. MEMS and NEMS necessitate sophisticated material characterization (MC) techniques.
- Availability and reliability are two sides of a coin. More reliable implies more available.
- The availability can be considered as the quantitative expression of reliability.
- Traceable calibration is a concept that ensures that all measurements made are accurate and valid.
- Sensor validation is taken to mean the generation of validity index of the sensed signal. Along with measured value a truthfulness value or more accurately the *validity index* (VI) is generated.
- Trimming is a process of removal of unusual portions prior to its final production.
- The structural imperfection is compensated through additional design considerations.
- The purpose of fabricating array configured MEMS devices is multifold. Besides the usefulness of array configuration for achieving a generic validated signal at the output of a sensor, the interest in array also lies with other prospects and viewpoints.
- Silicon Light Machines has developed a unique display method for the projection of imago-electric signal using its Grating Light Valve (GLV) principle based MEMS technology.
- DMD, a micromechanical silicon chip, is considered as a spatial light modulator (SLM).
- All physical, chemical and biological reactions occur with change in heat. Sensing mechanism has been designed to access the thermal characteristics of such reactions.
- Frequency selection is defined as selection of desired frequency out of many components. Frequency can be selected by using Frequency Selective Surface (FSS).
- The FSS functions as a spatial filter.
- FSS is a conducting sheet, understood as arrays of metallic patches which are periodically perforated with apertures.

Exercises

1. Define the term performance of a system.
2. Write down the important performance measuring parameters of a system? Precisely define these terms.
3. What do you mean by transient response of a system? Draw the typical output response of a system if the input is a unit step function. In this context, define the terms that are involved in describing the transient-response specification.
4. What do you mean by hysteresis? Mathematically what does it mean? Draw a typical hysteresis curve and explain its features. Does hysteresis affect performance? Give an example showing that hysteresis is also necessary in designing a control function.
5. Define the following terms in a hysteresis curve.
   (a) Retentivity (b) Coercivity (c) Opposite magnetic saturation

6. What is the meaning of stability of a system? If the input to a system is a step signal, pictorially show the different ways of expressing the instability within the system. From the figures explain what they actually express.
7. Define the meaning of reliability. Draw the reliability curve and categorize the regions within the curve. Discuss the characteristics of each region.
8. Enlist the advantages, which can be gained by implementing the reliability program in a manufacturing sector.
9. What important points are to be considered in order to achieve highly reliable products?
10. Define the term availability. Express availability in terms of reliability. In particular define MTBF, FIT and MTTR. Write down the mathematical expression of availability in terms of MTBF and MTTR.
11. What do you mean by traceable calibration? How can traceable calibration help in improving performance? What is validation index of a typical sensor? How can validation index be generated?
12. Discuss how compensation can be made in case of imperfect design?
13. What are the merits of designing arrayed MEMS system? Give some example of arrayed systems.
14. Write down the applications of the following arrayed system.
    (a) GLV technology
    (b) DLP technology
    (c) Calorimetric MEMS sensor array
    (d) FSS

References

Chapter 1

www.aspe.net
http://www.memsnet.org/mems/what-is.html
www.memsoptical.com/techinfo/memstut1.htm
http://www.aero.org/publications/helvajian/index.html

French, P. J. and P. M. Sarroz, "Surface versus bulk micromachining: The contest for suitable applications," *J. Micromech. Microeng.*, 8, 1998, pp. 45–53.

Lee, Sangwoo, Sangjun Park, and Dong-il Cho, "The Surface/Bulk Micromachining, SBM, Process: A New Method for Fabricating Released MEMS in Single Crystal Silicon," *J. of Microelectromechanical Systems*, 8/4, Dec. 1999, p 409.

Mahalik, N. P. "Mechatronics: Principles, Concepts and Applications, McGraw Hills, 2004.

Mahalik, N. P., Micromanufacturing & Nono technology, Springer–Verlog, Germany, 2006.

Maithripala, D. H. S., Jordan M. Berg and W. P. Dayawansa, "Capacitive Stabilization of an Electrostatic Actuator: An Output Feedback Viewpoint", *IEEE Proc. of the American Control Conf*. Denver, Colorado, June 4–6, 2003.

Mitchell, Jay and Jae Yoong Cho, *The Future of Wireless Devices*.

Mehregany, M. and S. Roy, Microfabrication Laboratory, C.W.R. Univ., Chapter 1: Introduction to MEMS, Published by the Aerospace Corporation.

Chapter 2

http://aveclafaux.freeservers.com/SU-8.html
Cheng-Hsien Liu, Micro-Electro-Mechanical Transducers, NTH Univ., Taiwan, 2004.
http://mx.nthu.edu.tw/~chhsliu/transducer/PME5230_Liu.html
http://www.elec.gla.ac.uk/groups/sim_centre/courses/diffusion/diff_5.html
www.SemiconFarEast.com.
http://www.memsnet.org/mems/beginner/deposition.html
http://www.semiconfareast.com/sputtering.htm

Rasmussen, John, William Bonivert, John Krafcik (AXSUN Technologies, Inc.), "High Aspect Ratio Metal MEMS (LIGA) Technologies for Rugged, Low-Cost Firetrain and Control Components", *INDIA 47th Annual Fuze Conf.*, April 10, 2003.

Ueno, H., M. Hosaka, Y. Zhang, O Tabata, S. Konishi and S. Sugiyama, "Study on Fabrication of High Aspect Ratio Microparts Using the LIGA Process", *IEEE Int. Symp. on Microelectronics and Human Science*, 1997.

Copyright © 2004 www.SemiconFarEast.com. All Rights Reserved.

Zou, Jun, Chang Liu, Jose Schutt-Aine, Jinghong Chen, and Sung-Mo Kang, "Development of a Wide Tuning Range MEMS Tunable Capacitor for Wireless Communication Systems", Copyright IEEE, 2000.

Allen, D. M., *The Principles and Practice of Photochemical Machining and Photoetching*, Adam Hilger, Bristol and Boston, 1986.

Allongue, P., V. Costa-Kieling, and H. Gerischer, "Etching of Silicon in NaOH Solutions", Part I and II, *J. Electrochem. Soc*., 40, 1009–1018, Part I, and 1018–1026, Part II, 1993.

Ammar, E. S. and T. J. Rodgers, UMOS Transistors on, 110, Silicon, *IEEE Trans. Electron Devices*, ED-27, pp. 907–914, 1980.

Bassous, E., H. H. Taub, and L. Kuhn, "Ink Jet Printing Nozzle Arrays Etched in Silicon", *Appl. Phys. Lett*., 31, pp. 135–137, 1977.

Boivin, L. P., "Thin Film Laser-to Fiber Coupler", *Appl. Opt*., 13, pp. 391–395, 1974.

Editorial, Transducers, Pressure and Temperature, Catalog, National Semiconductor, Sunnyvale, CA, 1974.

Editorial, Thermal Character Print Head, Texas Instruments, Austin, 1977.

Gad-el-Hak, M., *The MEMS Handbook* CRC press, 2002.

Greenwood, J. C., "Etched Silicon Vibrating Sensor", *J. Phys. E, Sci. Instrum*., 17, pp. 650–652, 1984.

Greenwood, J. C., Ethylene Diamine-cathechol-water mixture shows preferential etching of p-n junction, *J. Electrochem. Soc*., 116, pp. 1325–1326, 1969.

Hallas, C. E., "Electropolishing Silicon", *Solid State Technology*, 14, pp. 30–32, 1971.

Harris, T. W., "Chemical Milling", Clarendon Press, Oxford, 1976.

Hoffmeister, W., "Determination of the Etch Rate of Silicon in Buffered HF Using a 31 Si Tracer Method", *Int. J. Appl. Radiation and Isotopes*, 2, 139, 1969.

Hu, J. Z., L. D. Merkle, C. S. Menoni, and I. L. Spain, "Crystal Data for High-Pressure Phases of Silicon", *Phys. Rev. B*, 34, pp. 4679–4684, 1986.

Kaminsky, G., "Micromachining of Silicon Mechanical Structures", *J. Vac. Sci. Technol*., B3, pp. 1015–1024, 1985.

Kern, W. and C. A. Deckert, "Chemical Etching, in Thin Film Processes", Vossen, J. L. and Kern, W. (Eds.), Academic Press, Orlando, 1978.

Lee, J. B., Z. Chen, M. G. Allen, A. Rohatgi, and R. Arya, "A Miniaturized High-Voltage Solar Cell Array as an Electrostatic MEMS Power Supply", *J. Microelectromech. Syst*., 4, pp. 102–108, 1995.

Linde, H. and L. Austin, "Wet Silicon Etching with Aqueous Amine Gallates", *J. Electrochem. Soc*., 139, pp. 1170–1174, 1992.

Madou, M. J. and S. R. Morrison, *Chemical Sensing with Solid State Devices*, Academic Press, New York, 1989.

Middlehoek, S. and U. Dauderstadt, "Haben Mikrosensoren aus Silizium eine Zukunft?" *Technische Rundschau*, July, pp. 102–105, 1994.

Murarka, S. P. and T. F. J. Retajczyk, "Effect of Phosphorous Doping on Stress in Silicon and Polycrystalline Silicon", *J. Appl. Phys*., 54, pp. 2069–2072, 1983.

O'Neill, P., "A Monolithic Thermal Converter", *Hewlett-Packard J*., 31, pp. 12–13, 1980.

Palik, E. D., J. W. Faust, H. F. Gray, and R. F. Green, "Study of the Etch-Stop Mechanism in Silicon", *J. Electrochem. Soc.,* 129, pp. 2051–2059, 1982.

Peeters, E., "Process Development for 3D Silicon Microstructures, with Application to Mechanical Sensor Design", Ph.D. thesis, Catholic Univ. of Louvain, Belgium, 1994.

Petersen, K. E., "Silicon as a Mechanical Material", *Proc. IEEE*, 70, pp. 420–457, 1982.

Pfann, W. G., "Improvement of Semiconducting devices by elastic strain", *Solid State Electron*., 3, pp. 261–267, 1961.

Robbins, H. and B. Schwartz, *Chemical Etching of Silicon-II*. "The system HF, HNO_3, H_2O, and $HC_2C_3O_2$", *J. of Electrochem. Soc*., 107, pp. 108–111, 1960.

Rodgers, T. J., W. R. Hiltpold, B. Frederick, J. J. Barnes, F. B. Jenné, and J. D. Trotter, "VMOS Memory Technology", *IEEE J. Solid-State Circuits*, SC-12, 515–523, 1977.

Schwartz, B. and H. Robbins, "Chemical Etching of Silicon-IV, Etching Technology", *J. Electrochem. Soc*., 123, pp. 1903–1909, 1976.

Smith, C. S., "Piezoresistance Effect in Germanium and Silicon", *Phys. Rev*., 94, pp. 42–49, 1954.

Tuck, B., "Review—The Chemical Polishing of Semiconductors", *J. Mater. Sci*., 10, pp. 321–339, 1975.

Tufte, O. N., P. W. Chapman, and D. Long, "Silicon Diffused-element Piezoresistive Diaphragms", *J. Appl. Phys*., 33, pp. 3322, 1962.

Turner, D. R., "Electropolishing Silicon in Hydrofluoric Acid Solutions", *J. Electrochem. Soc*., 105, pp. 402–408, 1958.

Uhlir, A., "Electrolytic Shaping of Germanium and Silicon", *Bell Syst. Tech. J*., 35, pp. 333–347, 1956.

Waggener, H. A., R. C. Krageness, and A. L. Tyler, "Two-way Etch", *Electronics*, pp. 40, 274, 1967.

Craven, David "Photolithography Challenges for the Micromachining Industry", *BACUS Symp.* 96, Ultratech Stepper, San Jose, CA 95134.

Waits, Christopher M., Reza Ghodssi, and Madan Dubey, "Gray-Scale Lithography for MEMS Applications",

Dotson, Nathan A., Peter T. Kim and Andrew Mason, "Low Cost MEMS Processing Techniques", *Proc. of the 2004 ASEE/NCS Spring Conf*., April 1–3, 2004 Copyright 2004, American Society for Engineering Education.

Chiu, G. L.-T. and J. M. Shaw, Guest editors, "Optical lithography: Introduction, IBM", *J. of Research and Development*, Jan. 24, 1997.

"Micromachining and Microfabrication Process Technology VIII", John A. Yasaitis, Mary Ann Perez-Maher, Jean Michel Karam (Eds.), "SU-8 based deep X-ray lithography/LIGA", *Proc. of SPIE, Vol*. 4979, 2003, © 2003 SPIE.

Cheng, Yao Bor-Yuan Shew, Ching-Yao Lin, and Der-Hsin Wei "Deep X-ray Lithography Developed at SRRC", *Proc. Natl. Sci. Counc. ROC*, A, Vol. 23, No. 4, 1999. pp. 537–543.

Nallani, Arun Kumar, Sang Won Park, and Jeong Bong Lee, "Characterization of SU-8 as a resist for electron beam lithography", Micro/Nano Devices and Systems, MiNDS, Laboratory, Univ. of Texas at Dallas.

Michel B., A. Bernard, A. Bietsch, E. Delamarche, M. Geissler, D. Juncker, H. Kind, J.-P. Renault, H. Rothuizen, H. Schmid, P. Schmidt-Winkel, R. Stutz, and H. Wolf, "Printing meets lithography: Soft approaches to high-resolution patterning", *IBM J. of Research and Development*, Feb. 10, 2001.

Xia, Younan and George M. Whitesides, "Soft Lithography," *Annu. Rev. Mater. Sci*. 1998. 28:153–84.

Jian, Linke Yohannes M. Desta, Jost Goettert, Martin Bednarzik, Bernd Loechel, Jin Yoonyoung, Georg Aigeldinger, Varshni Singh, Gisela Ahrens, Gabi Gruetzner, Ralf Ruhmann, Reinhard Degen, "SU-8 based deep X-ray lithography/LIGA, Micromachining and Microfabrication Process Technology VIII", John A. Yasaitis, Mary Ann Perez-Maher, Jean Michel Karam, (Eds.) *Proc. of SPIE*, Vol. 4979, 2003.

Hafizovic, S., Y. Hid, O. Tahata and J.G. Korvink, "X3D: 3D X-Ray lithography and developement simulation for MEMS, Transducers '03", 12th Int. Conf. on Solid State Sensors, Actuators and Microsystems, Boston. June 8–12 2003, *IEEE* 1570.

Guckel, H., "High-Aspect-Ratio micromachining Via Deep X-Ray lithography", *Proc. of the IEEE*, 86/8, Aug. 1998, pp 1586–93.

Sugiyama, Susumu, Sommawan Khumpuang and Gaku Kawaguchi, "Plain-pattern to cross-section transfer", "PCT", "technique for deep X-ray lithography and applications", Inst. of physics publishing, *J. Micromech. Microeng*., 14, 2004.

Chase, J. Geoffrey, "Overview of Modern Lithography Techniques and a MEMS-Based Approach to High Throughput Rate", "Electron Beam Lithography", *J. of Intelligent Material Systems and Structures*, SAGE Publications, 12/12, pp 807-817, 2001.

Baborowski, J., R. Lanz, PD Dr. P. Muralt, N. Ledermann, Ceramics Laboratory, "MEMS: a Playground for New Thin Film Materials", Swiss Federal Inst. of Technology EPFL, Lausanne, Switzerland.

Fan, Shih-Kang and Chang-Jin Kim, "MEMS with thin-film aerogel," *IEEE*, 2001.

Hoffman D.M., P. Rangarajan, S.D. Athavale, D. J. Economou, Jia-Rui Liu, Z. Zheng, W-K Chu, "Plasma-enhanced chemical vapor deposition of silicon, germanium, and tin nitride thin films from metalorganic precursors," *J. Vac. Sci. Technol*. A 13/3, May/Jun 1995.

Eli´aš, P, I Kosti ǰc, J Šolt´ys and S Hasen¨ohrl, "Wet-etch bulk micromachining of, 100, InP substrates", Inst. of physics publishing *journal of Micromechanics and Microengeneering*, *J. Micromech. Microeng*., 14, pp. 2004, 1205–1214.

Oehrlein, G. S., M. F. Doemling, B. E. E. Kastenmeier, P. J. Matsuo, N. R. Rueger, M.Schaepkens, and T. E. F. M. Standaert, "Surface science issues in plasma etching," *IBM J. of Research and Development.*

Chu, Huai-Yuan and Weileun Fang "A Novel Convex Corner Compensation for Wet Anisotropic Etching on, 100, Silicon Wafer", *IEEE* 2004.

Schr¨ oder, H and E Obermeier, "A new model for Si{100} convex corner undercutting in anisotropic KOH etching", *J. Micromech. Microeng*. 10, 2000, 163–170.

Kloeck, Ben, Scott D. Collins, Nico F. DE Rooij, and Rosemary L. Smiyh, "Study of Electrochemical Etch-Stop for High-Precision Thickness Control of Silicon Membranes," *IEEE T. on Electron devices*, 36/4. April 1989.

Gregory, T. A. Kovacs, Nadim I. Maluf, and Kurt E. Petersen, "Bulk Micromachining of Silicon", *Proc. of IEEE*, 1998.

French, P.J.; Sarro P.M., "Surface versus bulk micromachining: the contest for suitable applications", *J. of Micromech. Microeng*., 1998, 8/2, pp. 45–53,9.

Lee, Sangwoo, Sangjun Park, and Dong-il Cho, "The Surface/Bulk Micromachining, SBM, Process: A New Method for Fabricating Released MEMS in Single Crystal Silicon", *J. of Microelectromechanical Systems*, 8/4, Dec. 1999, p 409.

Bustillo, James M., Roger T. Howe, and Richard S. Muller, "Surface Micromachining for Microelectromechanical Systems", *Proc. of the IEEE*, 86/8, Aug. 1998.

Chapter 4

http://www.cabaret.co.uk/
http://www.cabaret.co.uk/education/geneva.htm

Fan. L.S., Tai Y.C., Muller R.S., "Pin joints, gears, springs, cranks and other novel microstructures", *Tech. Dig., 4th Int. Conf. on Solid-state Actuators and Sensors*: pp. 849–852.

Clements, Deanne, Larry L. Howell, Nathan Masters, and Brent L. Weight, "Floating Pin Joints Fabricated from Two Layers of Polysilicon at the Micro Level".

Fan, L. Y. C. Tai, R. S. Muller, "Integrated Movable Micromechanical Structures for Sensors and Actuators", *IEEE Transactions on Electron Devices*, Vol. 15, No 6. June 1988.

Afridi, M., A. Hefner, D. Berning, C. Ellenwood, A. Varma, B. Jacob, S. Semancik, "MEMS-based embedded sensor virtual components for system-on-a-chip (SoC)", *Solid State Electronics*, Elsevier, 48 (2004) 1777–1781.

Ghantasala, M. K., L. Qin, D. K. Sood and R. B. Zmood, "TECHNICAL NOTE: Design and fabrication of a micromagnetic bearing", *Smart Mater. Struct*., 9 (2000) 235–240.

Smith, James H., Micromachined Sensor and Actuator Research at Sandia's Microelectronics Development Laboratory, Intelligent Micromachines Dept., Sandia National Laboratories, Sensors Expo, Philadelphia, Oct. 1996.

Mahalik, N. P., *Mechatronics; Principle, Concepts and Application*: McGraw-Hill Int. Edn., New York. 2004

Judy, Jack W., Richard S. Muller and Hans H. Zappe, "Magnetic Microactuation of Polysilicon Flexure Structures", Berkeley Sensor & Actuator Center, Univ. of California, Berkeley, Solid-State Sensor and Actuator Workshop, 1994.

Isermann, R. and U. Rabb., "Intelligent Actuators—Ways to Autonomous Actuating Systems", "*Symposium on Intelligent Components and Instruments for Control Application*, *SISICA-92*," Malaga, Spain, 1992.

Isermann, R, "Intelligent Actuators—Ways to Autonomous Actuating Systems", *Automatica*, Vol. 29, No. 5, 1993.

Isermann, R. "Integration of Fault Detection and Diagnosis Methods", *IFAC Symp. Safeprocess'94,* Helsinki, 1994.

Isermann, R. and P. Balle. "Trends in the Application of Model-based Fault Detection and Diagnosis of Technical Processes", *Proceeding, 13th IFAC Conference*, San Francisco, USA.

Chapter 5

http://documents.exfo.com/appnotes/AnoteBurleigh010-ang.pdf
http://www.piezomechanik.com/hp030530/pdfs/introduction.pdf

CNF, Cornell Nanofabrication Facility, Cornell Univ., Ithaca, NY 14853, http://www.research.cornell.edu/VPR/vpr.html, 1998.
http://microfluids.engin.brown.edu/Breuer_Papers/Conference.s/AIAA99-0606_Microphone.pdf
http://www.efunda.com/formulae/solid_mechanics/beams/casestudy_bc_cantilever.cfm

Tseng-Yang Hsu, Wen H. Hsieh, Yu-Chong Tai, Katsusi Furutani, "A thin film Teflon electret technology for microphone applications", California Institute of Technology, USA, and Toyota Technological Institute, Japan,

Neumann, J. J. and K. J. Gabriel, "A fully integrated CMOS-MEMS audio microphone", *12th Int. Conf. on Solid State Sensors, Actuators and Microsystems*, Boston, June 8-12, 2003.

Application note, Inchworm Motor Technology; Inchworm Motors Assure Precise Positioning, EXFO, 2003 EXFO Burleigh Products Group Inc., 7647 Main St. Fishers, Victor, NY 14564-8909, Printed in USA., Burleigh is a trademark and Inchworm is a registered trademark of EXFO Burleigh Products Group Inc.

Painter, Chris C. Student Member, IEEE, and Andrei M. Shkel, Associate Member, IEEE Active Structural Error Suppression in MEMS Vibratory Rate Integrating Gyroscopes, *IEEE Sensors J*., 3/5, Oct., 2003, p. 595.

Graff, J.W., E.F. Schubert, Flat free-standing silicon diaphragms using silicon-on-insulator wafers, *Sensors and Actuators*, Elsevier Science, 84, 2000, pp 276–279.

Su, Y, A. G. R. Evans and A. Brunnschweiler, "Micromachined silicon cantilever paddles with piezoresistive readout for flow sensing", *J. Micromech. Microeng*., 6, 1996, pp 69–72.

Keoschkerjan, R., H. Wurmus, "A novel microgripper with parallel movement of gripping arm", *Actuator* 2002, *8th Int. Conf. on New Actuators*, Germany, June 10–12, pp. 321–324, 2002.

Koganezawa, Shinji and Takeyori Hara, "Development of Shear-Mode Piezoelectric Microactuator for Precise Head Positioning, Fujitsu", *Magazine*, 37, 2, 2001.

Ounaies, Zoubeida, Joycelyn S. Harrison and Rich J. Silcox, "Piezoelectric Materials for Sensor and Actuator Applications at NASA LaRC", *ICASE Research quarterly*, 8/2, June 1999.

Wang, X., B. Li, S. Lee, Y. Sun, H.T. Roman, K. Chin and K.R. Farmer, "A New Method to Design Pressure Sensor Diaphragm", Nanotech 2004, *Technical Proc. of the 2004 NSTI Nanotechnology Conf. and Trade Show*, Vol. 1.

Chan-Park, M. B., Neo, W. K., "Ultraviolet embossing for patterning high aspect ratio polymeric microstructures", *Microsystem Technologies*, 9/6-7–501–506, 2004.

Timoshenko, S. and S. Woinosky-Krieger, *Theory of Plates and Shells*, McGraw-Hill Classic Textbook Reissue, 1987.

Schellin, R., G. Hess, W. Kühnel, C. Thielemann, D. Trost, J. Wacker, and R. Steinmann, *Sensors and Actuators A*, 41–42, pp. 287–292, 1994.

Qiu* Jin, Jeffrey H. Lang, Alexander H. Slocum, "A Centrally-Clamped Parallel-Beam Bistable MEMS Mechanism", Massachusetts Inst. of Technology.

Belegundu, A.D. and Chandrupatla, T.R. *Optimization Concepts and Applications in Engineering*, pp. 272 – 276, Prentice Hall Publishing, NJ, 1999.

Burleigh, "Inchworm nanopositioning systems brochure", Burleigh Instruments Inc., Fishers, USA, 1995.

Frank, J., "Design and Development of Piezoelectric Motors", Ph.D. Thesis, Penn State Univ., 2001.

Frank, J., Mockensturm, E., Chen, W., Koopmann, G.H. and Lesieutre, G.A., "Roller Wedgeworm: A piezoelectrically-driven rotary motor", *10th Int. Conf. on Adaptive Structures and Technologies*, Paris, 1999.

Galante, "Design and Fabrication of a High-force'Linear'Piezoceramic Actuator", M.S. Thesis, Penn State Univ., 1997.

Galante, T., Frank J., Bernard J., Chen W., Lesieutre G.A. and Koopmann G.H., Design, "Modeling and Performance of a High Force Piezoelectric Inchworm Actuator", In: *Proc. of the SPIE 1998 Conf. on Smart Materials and Structures*, March 1998.

Koopmann, G.H., Lesieutr, G.A., Fran, J.E. and Chen, W., "Design and Performance of a Linear Piezoelectric Wedgeworm Actuator", *NATO Advanced Research Workshop-Predeal*, Romania, May 1999.

Lee, S.K. and Esachi, M., "Design of the Electrostatic Linear Microactuator Based on the Inchworm Motion, *Mechatronics*, 5/ 8: 9653–9972, 1995.

Meisner, J.E. and Teter, J.P., "Piezoelectric/Magnetostrictive Resonant Inchworm Motor", *SPIE*, 2190: 520–527, 1994.

Pandell, and Garcia, E., "Design of a Piezoelectric Caterpillar Motor", In: *Proc. of the ASME Aerospace Division*, AD-52: 627–648, 1996.

Armitage, A., N.R. Scales, P.J. Hicks, P.A. Payne, Q.X. Chen, and J.V. Hatfield, "An Integrated Array Transducer Receiver for Ultrasound Imaging", *Sensors and Actuators A*, Vol. 47, pp. 540–544, 1995.

Baltes, H., "Future of IC Microtransducers", *Sensors and Actuators A*, Vol. 56, pp. 179–192, 1996.

Borner, M., S. zur Horst-Meyer, M. Murphy, H. Munch, W. Schomburgand M. Vitt, "Ultrasonic Measurements with Micromembranes", *Sensors and Actuators A*, Vol. 46, pp. 62–65, 1995.

Davidsen, R., R. Jensen and S. Smith, "Two-Dimensional Random Arrays for Real Time Volumetric Imaging", *Ultrasonic Imaging*, Vol. 16, pp. 143–163, 1994.

Sheplak, Mark, John M. Seinery, Kenneth S. Breuerz Martin A. Schmidt, "A MEMS Microphone for Aeroacoustics Measurement", American Inst. of Aeronautics and Astronautics, Inc. 1999.

Acar, Cenk and Andrei M. Shkel, "Structural design and experimental characterization of torsional micromachined gyroscopes with non-resonant drive mode", *J. of Micromech. Microeng.*, 14, pp 15–25–2004.

Sze, S.M., *Semiconductor Sensors*, Wiley-Interscience, New York, 161–179, 1994.

Stibitz, G.R., *Incremental Feed Mechanisms*, Patent No. 3,138,749, 1964.

Chee, Y. H., C. Tsang, "Micro-Robot Leg Design for Smart dust using Improved Inchworm Motor", Dept. of Electrical Engineering and Comp. Sc., Cory Hall, Univ. of California at Berkeley.

Shearwood, C. and C.B. Williams, P.H. Mellor, and R.B. Yates, "Micromachined Rotating Gyroscope", Seoul National Univ., Korea.

Alper, S.E. and T. Akin, "A Planar Gyroscope using a Standard Surface Micromachining Process", *14th European Conf. on Solid-State Transducers*, Eurosensors XIV, pp. 387–390, Copenhagen, Aug. 27–30, 2000.

Acar, Cenk and Andrei M. Shkel, "Structural design and experimental characterization of torsional micromachined gyroscopes with non-resonant drive mode", *J. of Micromech. Microeng.*, 14, 2004, pp 15–25.

Popova, I. V., A. M. Lestev, E. N. Pyatyshev, and M. S. Lur'e "Micromechanical Gyroscopes and Areas of Their Application", GIROOPTICA Ltd., St Petersburg, Russia.

Tung, Steve, Position Paper: "An Overview of MEMS Inertial Sensors", Dept. of Mech. Engineering, Univ. of Arkansas, Arkansas.

Chapter 6

Peltier coolers, Victor Rudometov, Eugene Rudometov at http://www.digit-life.com/articles/peltiercoolers/

http://herkules.oulu.fi/isbn9514252217/html/x317.html

http://www.cs.ualberta.ca/~database/MEMS/sma_mems/sma.html

Rudometov, E. and V. Rudometov, *PC Overclocking, Optimization, and Tuning*, book by ISBN 1-58450-079-4, bhv PUBLISHING HOUSE,

Lin, Qiao, Fukang Jiang, Xuan-Qi Wang, Yong Xu, Zhigang Han, Yu-Chong Tai, James Lew and Chih-Ming Ho, "Experiments and simulations of MEMS thermal sensors for wall shear-stress measurements in aerodynamic control applications", *J. of Micromech. Microeng.*, 14, 2004.

Yan, Dong, MS thesis, "Mechanical Design and Modeling of MEMS Thermal Actuators for RF Applications", Univ. of Waterloo, 2002.

Yan, Dong, Amir Khajepour and Raafat Mansour, "Modeling of two-hot-arm horizontal thermal actuator", *J. of Micromech. Microeng.*, Inst. of Physics Publishing, 13, 2003.

Yan, Dong, Amir Khajepour and Raafat Mansour, "Design and modeling of a MEMS bidirectional vertical thermal actuator", *J. of Micromech. Microeng.*, Inst. of Physics Publishing, 14, 2004.

van Oudheusden, B.W. Ph.D. "thesis, Integrated Silicon Flow Sensors", Delft Univ. of Technology, 184 pp., 1989.

Snyder, G. Jeffrey, James R. Lim, Chen-Kuo Huang and Jean-Pierre Fleurial, *Nature Materials*, Vol. 2, pp. 528 to 531, "Thermoelectric microdevice fabricated by a MEMS-like electrochemical process", 2003.

Pagonis, D. N., G. Kaltsas and A. G. Nassiopoulou, "Fabrication and testing of an integrated thermal flow sensor employing thermal isolation by a porous silicon membrane over an air cavity", *J. of Micromech. Microeng.*, Inst. of Physics Publishing, 14, 793–797, 2004.

Kaltsas, G. and Nassiopoulou A. G., "Novel C-MOS compatible monolithic silicon gas flow sensor with porous silicon thermal isolation", *Sensors Actuators A*, 76, 133, 1999.

Kovacs, G.T.A., *Micromachined Transducers Sourcebook,* WCB/McGraw-Hill, 1998.

Putten, F. P. van, *Electronic Measurement Systems,* IOP Publishing, 1996.

Bentley, J. P., *Principles of Measurement Systems,* 3rd Edn., Longman Group Ltd., 1995.

Carstens, J. R., *Electrical Sensors and Transducers,* Regents/Prentice Hall, 1993.

Hauptmann, P., *Sensors, Principles & Applications,* Prentice Hall, 1993.

DeVoe Don, L., "Thermal Issues in MEMS and Microscale Systems", *IEEE T. on Components and Packaging,* 25/4, Dec. 2003.

Ming, H. Wu, "Fabrication of Nitinol Materials and Components", *Proc. of the Int. Conf. on Shape Memory and Superelastic Technologies,* Kunming, China, pp. 285–292, 2001.

Woudenberg, T. and M. Albin, Perkin-Elmer Applied Biosystems.

Chui, Benjamin W., Timothy D. Stowe, Yongho Sungtaek Ju, Kenneth E. Goodson, Thomas W. Kenny, H. Jonathon Mamin, Bruce D. Terris, Robert P. Ried, and Daniel Rugar, "Low-Stiffness Silicon Cantilevers with Integrated Heaters and Piezoresistive Sensors for High-Density AFM Thermomechanical Data Storage", *J. of MEMS,* 7/1, March 1998.

Mo-Huang, Li, Julius J. Wu, and Yogesh B. Gianchandani, Surface Micromachined Polyimide Scanning Thermocouple Probes, *J. of MEMS,* 10/1, March 2001.

Venkataraman, C., E. M. Chow, T. W. Kenny, T. N. Louis, N. Cattafesta, B. V. Sankar and M. Sheplak, "Thermoelestically actuated acoustic proximity sensor with integrated through-wafer interconnects", Solid-State Sensor, Actuator and Microsystems Workshop Hilton Head Island, South Carolina, June 2–6, 2002.

Venkataraman, C., B. V. Sankar, L. N. Cattafesta, T. Nishida, and M. Sheplak, "An analytical model for the thermoelastic actuation of composite diaphragms", *12th Int. Conf. on Solid State Sensors, Actuators and Microsystems*, Boston, June 8–12, 2003.

Chang, D. T., D. M. Chen, F. H. Lin, W. J. Kaiser, O. M. Stafsudd, *CMOS Integrated Infrared Sensors*, Electrical Engineering Dept., Univ. of California, Los Angeles.

Ho, Jyh-Jier, Y. K. Fang, W. J. Lee, F. Y. Chen, W. T. Hsieh, S. F. Ting, M. S. Ju, S. B. Huang, Kun-Hsien Wu, and C. Y. Chen, "The Dynamic Response Analysis of a Pyroelectric Thin-Film Infrared Sensor with Thermal Isolation Improvement Structure", IEEE T. on Electron Devices, 46/12, Dec. 1999.

Jin Qiu, PhD thesis, "An Electrothermally-Actuated Bistable MEMS Relay for Power Applications", Massachusetts Inst. of Technology, 2003.

Ellis Meng and Yu-Chong Tai, "A parylene MEMS flow sensing array", *Technical Dig., 12th Int. Conf. on Solid-State Sensors, Actuators and Microsystems, Transducers '03*,, Boston, pp. 686-689, 2003

Luo, J. K., A. J. Flewitt, S. M. Spearing, N. A. Fleck and W. I. Milne, "Three types of planar structure microspring electro-thermal actuators with insulating beam constraints", *J. Micromech. Microeng*. 15, 2005, 1527–1535

Chapter 7

http://www.phys.hawaii.edu/~teb/optics/java/slitdiffr/

http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/sinslit.html#c1

Jianglong Zhang, Y. C. Lee, Victor M. Bright and *John Neff, Digitally Positioned micromirror for open-loop control applications, http://mems.colorado.edu/c1.res.pub/c2.ftp/jl_MEMS2002_final.pdf

http://www.edmundoptics.com/TechSupport/DisplayArticle.cfm?articleid=284#Optics: Beamsplitters

http://www.elecdesign.com/Articles/Index.cfm?ArticleID=5942

http://www.memsoptical.com/prodserv/products/finepitch.htm

http://www-leti.cea.fr/commun/europe/MOEMS/moems.htm.

www-bsac.eecs.berkeley.edu/ ~mattlast/research/moems99final.pdf

www-bsac.eecs.berkeley.edu/ ~mattlast/research/moems99final.pdf

Hsiharng Yang, Ching-Kong Chao, Che-Ping Lin and Sheng-Chih Shen 3 "Micro-ball lens array modeling and fabrication using thermal reflow in two polymer layers", *J. of Micromech. Microeng.,* 14, pp 277–282, 2004.

Feather, G.A. Monk, D.W. "The digital micromirror device for projection display", *Proceeding, Seventh Annual IEEE Int. Conf. on Wafer Scale Integration*, 1995.

Hocker, G. B., D. Youngner, E. Deutsch, S. Senturia, M. Butler, M. Sinclair, and A. J. Ricco, "The polychromator: A programmable MEMS diffraction grating for synthetic spectra".

Kuhn, M., C. Kaufmann, M. Flaspöhler, T. Wächtler, T. Gessner, A. Hübler, "Microactuator with Diffraction Grating, Inst. for Print and Media Technology", Chemnitz Univ. of Technology, Germany, Center for Microtechnologies, Chemnitz Univ. of Technology, Germany

Fourguette, D., D. Modarress, F. Taugwalder, D. Wilson1, M. Koochesfahani, M. Gharib2 Miniature and MOEMS Flow Sensors.

Allan, Roger, "3D-MEMS based optical switch", *Electronics Design,* Oct. 27, 2003.

Lee, Shi-Sheng, Long-Sun Huang, Chang-Jin Kim and Ming C. Wu, "2×2 MEMS fiber optic switches with silicon sub-mount for low-cost packaging", Electrical Engineering Dept., UCLA.

Bains, Sunny, "Light Constructions", OE Reports, June 1997.

Bloom, D.M., "The Grating Light Valve: revolutionizing display technology", Formerly Echelle, Inc.

Apte, R., F. Sandejas, W. Banyai and D. Bloom, "Grating Light Valves for High Resolution Displays", *Solid State Sensors and Actuators Workshop,* June 1994.

Born, M. and E. Wolfe, *Principles of Optics*, Pergamon Press, New York, 1959.

Goodman, J., *Introduction to Fourier Optics*, McGraw-Hill Book Company, San Francisco, pp. 61–63, 1968.

Veldkamp, W., G. Swanson, S. Gaither, C.-L. Chen and T. Osborne, "Binary Optics: a Diffraction Analysis", MIT Lincoln Laboratory, Aug. 23, 1989.

Solgaard, O., "Integrated Semiconductor Light Modulators for Fiber-Optic and Display Applications", Ph.D. Thesis, Stanford Univ., 1992.

Trisnadi, Jahja I., Clinton B. Carlisle and Robert Monteverde, "Overview and applications of Grating Light Valve^{TM} based optical write engines for high-speed digital imaging", *Photonics West 2004—Micromachining and Microfabrication Symp.,* CA, USA, Jan. 26, 2004.

Koyama, F., "MEMS and Hollow Optical Waveguides for Widely Tunable Devices EECS" 298-5 Optical MEMS Seminar, Microsystems Research Center, Tokyo Inst. of Technology, Japan.

Shin, W. and K. Oh, "Novel micro-optical waveguide on microactuating platform for re-configurable wavelength selective optical switch", *Optic Express,* Optical Society of America, Vol. 12, No. 19, Sept. 2004.

Glebov, Alexei L., Lidu Huang, Shinegori Aoki, Michael G. Lee, Kishio Yokouchi, "Two-dimensional microlens arrays in silica-onsilicon planar lightwave circuit technology", JM^3 2,4, 309–318, *Society of Photo-Optical Instrumentation Engineers,* Oct. 2003.

Sche, S. and M.L. Wu, "Technology Impact to the Competitiveness of Fiber Optic Communication Modules: MicroOptoElectroMechanicalSystems", *The 13th Annual Wireless & Optical Communication Conf.*, 2004.

Matthew Last, KSJ Pister, "2-DOF Actuated Micromirror Designed for Large DC Deflection", MOEMS '99, Mainz, Germany, Aug 1999.

Chapter 8

Cyrus Daboo, http://homer.phy.cam.ac.uk/Members/Cyrus_Daboo/MOKE/MOKE.html

Takeyama Hiroshi, Senoo Kazutaka, Takashi Otuki, Mitsuteru Kimura, Double Injection Type Magnetodiode formed on an SOI substrate, http://www.iee.or.jp/honbu/back_number/J./index_back_number/2002/2002_05e_08.pdf

http://lmis3.epfl.ch/research/old/high3d/

Eyre Beverley, Kristofer, S. J. Pister, Micromechanical Resonant Magnetic Sensor in Standard CMOS, *IEEE Electron Device Letters*, 19/12, Dec. 1998, http://www.ee.ucla.edu/~eyre/res/t97.pdf.

http://www.wtec.org/loyola/mems/toc.htm

William S. Trimmer, Ed, "*Micromechanics and MEMS: Classic and Seminal Papers to 1990",* Wiley-IEEE Press, Jan. 1997.

Gibbs1, M. R. J. E. W. Hill and P. J. Wright, Magnetic materials for MEMS applications, Inst. of Physics Publishing, *J. of Physics D: Applied Physics,* 37, 2004.

Judy1, Jack W. and Nosang Myung2, "Magnetic Materials for MEMS, Electrical Engineering1", *Chemical Engineering*2, Univ. of California, Los Angeles, CA, USA.

Grimes Craig A. Mahaveer K. Jain, R. S. Singh, Qingyun Cai, Andrew Mason, Kenichi Takahata, Yogesh Gianchandani, "Magnetoelastic microsensors for environmental monitoring", *Dig., IEEE Int. MEMS-01 Conf.*, Interlaken, Switzerland, Jan. 2001.

Il-seok Son, Amit Lal, "A remotely actuated magnetic actuator for microsurgery with piezoresistive feedback", *Electrical and Comp. Engineering,* Univ. of Wisconsin at Madison.

Wright, John A.,Yu-Chong Tai and Shih-Chia Chang, A Large-Force, Fully-Integrated MEMS Magnetic Actuator, Technical Dig.: 1997, *Int. Conf. on Solid-State Sensors and Actuators: Transducers '97*, Vol. 2, Chicago, IL, pp. 793–796, June 1997.

Jin, Cho Hyoung, PhD Thesis, Micromachined Permanent Magnets and Their MEMS Applications, Univ. of Cincinnati.

Moreland John, "Micromechanical instruments for ferromagnetic measurements", *J. of Physics,* Inst. of Physics Publishing, 36, R39–R51, 2003.

Ferris, S A, J M Ivison and D Walker, "The magnetoresistor as a displacement transducer element", *J. Phys. E: Sci. Instrum.,* 3, 639–642, 1970.

Lenz, James E., *A review of magnetic sensors, IEEE Proc*., ISSN 0018-9219, vol. 78, p. 973–989, June 1990.

Michael, J. Caruso Dr. Carl, H. Smith, Tamara Bratland, Robert Schneider, *A New Perspective on Magnetic Field Sensing.*

Lee Seung-Ki, Kwang-Hoon Oh, June-Koo Rhee, Kuk-Jin Jhun, Min-Koo Han, "A novel magnetotransistor based on the drift current in the emitter", *TRANSDUCERS '91, Int. Conf. on Solid-State Sensors and Actuators*, 1991.

Sung Guo-Ming, "Interaction between Magnetoresistor and Magnetotransistor in the Longitudinal and Folded Vertical Hall Devices", *IEEE Sensors J.*, 4/6, pp. 749, Dec., 2004.

Michael, J. Caruso, Lucky S. Withanawasam, *Vehicle Detection and Compass Applications using AMR Magnetic Sensors,* Honeywell, Plymouth, MN, USA.

Metz, Matthias and Henry Baltes, Senior Member, "Offset in CMOS Magnetotransistors—Part I: Analysis of Causes", *IEEE T. on Electron Devices,* 48/9, Sept. 2001.

Schott Ch., H. Blanchard, R. S. Popovic, R. Racz and J. Hrejsa, "High Accuracy Analog Hall Probe", *Conf. on Precision Electromagnetic Measurement*, Braunschweig, Germany, June 17-20, 1996.

Schott Ch., D. Manic, and R. S. Popovic, "Microsystem for High Accuracy 3D Magnetic Field Measurement, Eurosensors XI, Warsaw, Poland, Sept. 1997.

Karl, W. J., Powell, A. L., Watts, R., Gibbs, M. R. J. and Whitehouse, C. R. 2000 *Sensors Actuators*, 81, 137.

M. R. J. Gibbs1, E. W. Hill2 and P. J. Wright, Magnetic materials for MEMS applications, *J. of Physics D: Appl. Phys.* 37, 2004.

Jun Shen, Chandler, Ariz., Charles Wheeler, "MEMS switch uses magnetic actuation", Microlab Inc., July 12, 2001.

Cho Hyoung J. and Chong H. Ahn, A Bidirectional Magnetic Microactuator Using Electroplated Permanent Magnet Arrays, *J. of Microelectromechanical Systems,* 11/1, Feb. 2002.

Jack W. Judy, Richard S. Muller, and Hans H. Zappe, "Magnetic Microactuation of Polysilicon Flexure Structures", *J. of Microelectromechanical Systems,* 4/4, Dec. 1995.

Jun Yan, Selso Luanava, and Vincenzo Casasanta, "Magnetic Actuation for MEMS Scanners for Retinal Scanning Displays", *SPIE Proc. 4985, MOEMS Display and Imaging Systems*, ISBN 0-8194-4785-4.

Laure K. Lagorce, Oliver Brand, "Magnetic Microactuators Based on Polymer Magnets", *IEEE J. of Microelectromechanical Systems,* 8/1, March 1999.

Il-seok Son, Amit Lal, "A remotely actuated magnetic *actuated magnetic* actuator for microsurgery with piezoelectric feedback", *Electrical and Comp. Engineering,* Univ. of Wisconsin at Madison.

Chang Liu, Thomas Tsao, Yu-Chong Tai, Wenheng Liu, Peter H. Will, Chih-Ming Ho, "A Micromachined Permalloy Magnetic Actuator Array for Micro-Robotics Assembly Systems", *Int. Conf. of Solid State Sensors and Actuators,* Stockholm, Sweden, June 1995.

Liu, Chang, Member, IEEE, and Yong W. Yi "Micromachined Magnetic Actuators using Electroplated Permalloy", *IEEE T. on Magnetics,* 35/3, May 1999.

Jennifer W. L. Zhou, Ho-Yin Chan, Tony K. H. To, King W. C. Lai, and Wen J. Li, "Polymer MEMS Actuators for Underwater Micromanipulation", *IEEE/ASME T. on Mechatronics,* 9/2, June 2004.

Miller, R. A.,Y. C. Tai, G. Xu, J. Bartha, F. Lin, "An electromagnetic MEMS 2×2 Fiber Optic Bypass Switch", California Inst. of Technology and Physical Optic Corporation, USA.

Griffin, John Linwood, Steven W. Schlosser, Gregory R. Ganger, David F. Nagle, "Modeling and Performance of MEMS-Based Storage Devices", *Proc. of ACM SIGMETRICS 2000,* Santa Clara, California, June 17–21, 2000.

Mustafa Uysal, Arif Merchant, Guillermo A. Alvarez, Using MEMS-based storage in disk arrays, Appears in the *Proc. of the 2nd USENIX Conf. on File and Storage Technologies,* San Francisco, CA, March 2003.

Pantazi, A., MA Lantz, G. Cherubini, H. Pozidis and E. Eleftheriou *A servo mechanism for a micro-electromechanical-system-based scanning-probe data storage device, Nanotechnology*, Inst. of Physics Publishing, 15, 2004.

Chapter 9

Jinghong Chen, Jun Zou, Chang Liu and Sung-Mo, Steve, Kang, "Development of MEMS Vertical Planar Coil Inductors through Plastic Deformation Magnetic Assembly", PDMA, available at http://www.comppub.com/publ/MSM2002/287.pdf, Agere Systems, Holmdel, NJ 07733.

"RF MEMS Switches—Description of Technology", NASA electronics parts and packaging program, available at http://nepp.nasa.gov/index_nasa.cfm/813/_edn3.

El-Khoury, George K. and Mohammad A. Hotait, "SMA Actuated RF MEMS Switch", Dept. of Mech. Engineering, American Univ. of Beirut.

Beirut- Lebanon, available at http://webfea-lb.fea.aub.edu.lb/Proc./2004/SRC-ME-10.pdf

www.rfic.co.uk, Phase shifter design tutorial

Hector J De. Los Santos, Randy J. Richards, MEMS for RF/Microwave wireless applications: The nexe wave – Part II, Coventor Inc. Irvine, CA, also published in *Microwave J.,* July 2001 issue.

Univ. of Illinois at Urbana-Champaign, Urbana, Univ. of California.

Imed Zine-El-Abidine, Michal Okoniewski and John G. McRory, "A New Class of Tunable RF MEMS Inductors", *Proc. of the Int. Conf. on MEMS, NANO and Smart Systems, ICMENS'03*, 0-7695-1947-4/03 $17.00 © 2003 IEEE

Rebeiz Gabriel M., *RF MEMS: Theory, Design and Technology,* Wiley-Interscience, 2003.

Yong-Kyu Yoon, Jin-Woo Park, and Mark G. Allen, "RF MEMS based on epoxy-core conductors", Georgia Inst. of Tech.: http://mems.mirc.gatech.edu/msmawebsite/Projects/project_list_files/Tall_inductor.pdf

Technology Report, From March 2003 High Frequency Electronics, Copyright © 2003, Summit Technical Media, LLC: http://www.highfrequencyelectronics.com/Archives/Mar03/HFE0303_TechRpt.pdf

McCorquodale, Michael S., Mei Kim Ding, and Richard B. Brown, "A CMOS Voltage-to-Frequency Linearizing Preprocessor for Parallel Plate RF MEMS Varactors", *2003 IEEE MTT-S Dig.,* IFTU-54,

Zou, Jun, Chang Liu, Jose E. Schutt-Aine, "Development of a Wide-Tuning-Range Two-Parallel-Plate Tunable Capacitor for Integrated Wireless Communication Systems", *Int J RF and Microwave CAE* 11: 322_329, John Wiley & Sons, Inc. 2001.

Sankaranarayanan, Janakiram G., Manas Behera Narayan Aluru Kartikeya Mayaram, "Accuracy Issues in a High-LevelModel forMEMS Varactors". This work is supported in part by NSF grant CCR – 0121616.

Barbastathis, George, Gregory N. Nielson, Peter Vandre, Brad Smith, Mike Shirk, "Ultra-fast Low Actuation Voltage RF MEMS Switch", Despande Center for Technological Innovation, MIT, April 8, 2004.

Klaasse, G., R. Puers and H.A.C. Tilmans, "Piezoelectric versus electrostatic actuation for a capacitive RF-MEMS switch", IMEC, Kapeldreef 75, 3001 Heverlee, Belgium, K.U.Leuven, Kasteelpark Aremberg 10, 3001 Heverlee, Belgium.

Robeiz, Gabriel M., *RF MEMS – Theory, Design and Technology,* Wiley, 2003.

Bircumshawa, c, Brian, Gang Liub, Hideki Takeuchib, Tsu-Jae Kingb, Roger Howea,b,c, Oliver O'Reillya, Albert Pisano, "The radial bulk annular resonator: Towards a 50 micron RF MEMS filter", *TRANSDUCERS '03, 12th Int. Conf. on Solid State Sensors, Actuators and Microsystems*, Boston, June 8–12, 2003.

Campbell, Nick "Electromagnetic Modeling of RF MEMS Devices", "Case Study—Modeling and Analysis of a MEMS LC Filter", Ansoft Corporation Europe, published in TechOnLine.

Strohm, Karl M., Franz Josef Schmückle, Bernd Schauwecker, Johann-Friedrich Luy, Wolfgang Heinrich, "Silicon Micromachined RF MEMS Resonators", *IEEE MTT-S CDROM*, 2002.

Berger, Jill D., Yongwei Zhang, John D. Grade, Howard Lee, Stephen Hrinya, Hal Jerman, Al Fennema, Alex Tselikov and Doug Anthon, "Widely tunable external cavity diode laser using a MEMS electrostatic rotary actuator", Iolon, Inc., 1870 Lundy Ave., San Jose, California 95131.

Liu, Yu, Andrea Borgioli, Amit S. Nagra, and Robert A. York, "K-Band 3-Bit Low-Loss Distributed MEMS Phase Shifter", *IEEE Microwave and guided wave letters,* 10/10, Oct. 2000, pp. 415.

Ji, T S, K J Vinoy and V K Varadan, "Distributed MEMS phase shifters by microstereolithography on silicon substrates for microwave and millimeter wave applications", *Smart Mater. Structure.* 10, 2001, Inst. of Physics Publishing, pp 1224–1229.

Chapter 10

Bains, Sunny, "Simple microfluidic system tunes fiber properties", *LaserFocusWorld,* available at http://lfw.pennnet.com/Articles/Article_Display.cfm?Section=ARCHI&ARTICLE_ID=174567&VERSION_NUM=1&p=12

http://en.wikipedia.org/wiki/Electrowetting

http://stommel.tamu.edu/~baum/reid/book1/book/node56.html

http://hyperphysics.phy-astr.gsu.edu/hbase/surten2.html#c1

http://www.sali.freeservers.com/engineering/cfd/

http://en.wikipedia.org/wiki/Navier-Stokes_equations

Chiou, Pei Yu, Hyejin Moon, Hiroshi Toshiyoshi, Chang-Jin Kim, Ming C. Wu, "Light actuation of liquid by optoelectrowetting", *Sensors and Actuators A*, 104, 2003, 222–228.

Yun, Kwang-Seok, Il-Joo Cho, Jong-Uk Bu1, Geun-Ho Kim1, Young-Sam Jeon1, Chang-Jin, CJ, Kim2, and Euisik Yoon, "A Micropump Driven by Continuous Electrowetting Actuation for Low Voltage and Low Power Operations", Dept. of Electrical Eng. And Computer Sci., Korea Advanced Inst. of Science and Technology, KAIST, 373-1 Kusong-dong, Yusong-gu, Taejon 305–701, Korea 1LG Electronics Inst. of Technology, LG Elite, Seoul, Korea, 2Mechanical and Aerospace Eng. Dept., Univ. of California, Los Angeles, CA 90095-1597, USA.

Xu, Tian-Bing, Ji Su and Qiming Zhang, "Theoretical Evaluation of Electroactive Polymer Based Micropump Diaphragm for Air Flow Control", SPIE 11th Annual Int. Symp. on Smart Structures and Materials, San Diego, California, March 14–18, 2004,

Martynova, Larisa, Laurie E. Locascio, Michael Gaitan, Gary W. Kramer, Richard G. Christensen, and William A. MacCrehan, "Fabrication of Plastic Microfluid Channels by Imprinting Methods", Analytical Chemistry Division and Semiconductor Electronics Division, National Inst. of Standards and Technology, Gaithersburg, Maryland 20899, *Anal. Chem.* 1997, 69, 4783-4789.

http://www.uni-ulm.de/uni/fak/natwis/angphys/deutsch/projektgruppen/mugele/elektrowetting_eng.html

Chiou, Pei Yu, Zehao Chang, Ming C. Wu, *Light actuated microfluidic devices,* Dept. of Electrical Engineering, Univ. of California at Los Angeles, 420 Westwood Plaza, Los Angeles, CA 90095-1594, USA.

Ahmed R., D. Hsu, C. Bailey, and T.B. Jones, "Dispensing picoliter droplet using dielectrophoretic, DEP, micro-actuation", *1st Int. Conf. on Microchannels and Minichannels*, April 24–25, 2003, New York, USA.

Ramos, H. Morgan, N. G. Green, and A. Castellanos, "AC electrokinetics: a review of forces in microelectrode structures", *J. Phys D: Appl. Phys.,* vol. 31, pp. 2338-2353, 1998.

Green N. G., A. Ramos, A. González, A. Castellanos, and H.Morgan, "Electrothermally induced fluid flow on microelectrodes", *J. of Electrostatics,* vol. 53, pp. 71–87, 2001.

Green, N. G., A. Ramos, A. González, A. Castellanos, and H. Morgan, "Electric field induced fluid flow on microelectrodes: the effect of illumination", *J. Phys. D: Appl. Phys*., vol. 33, pp. L13–L17, 2000.

Bart, S. F., L. S. Tavrow, M. Mehregany, and J. H. Lang, "Microfabricated electrohydrodynamic pumps", *Sensors and Actuators,* vol. A21–A23, pp. 193–197, 1990.

Fuhr, G., R. Hagedorn, T. Müller, W. Benecke, and B. Wagner, "Microfabricated electrohydrodynamic, EHD, pumps for liquids of higher conductivity", *J. Microelectromech. Syst.,* vol. 1, pp. 141–146, 1992.

[8] Fuhr, G., T. Schnelle, and B. Wagner, "Traveling wave-driven microfabricated electrohydrodynamic pumps for liquids", *J. of Micromech. and Microeng.,* 4, pp. 217–226, 1994.

Bohl, Steger R. B., A. Neurath, S. Messner, H. Sandmaier, R. Zengerle, P. Koltay, "A Highly Parallel Nanoliter Dispensing System Fabricated by High-speed Micromilling of Polymers", Univ. of Freiburg and HSG-IMIT, Germany.

Meng, Ellis and Yu-Chong Tai, "A paryleme MEMS flow sensing array," Dept. of Electrical Engineering, Caltech Micromachining Lab., California Inst. of Technology, pp.136–93, Pasadena, CA 91126, USA, *Technical Dig., The 12th Int. Conf. on Solid-State Sensors, Actuators and Microsystems, Transducers '03*, Boston, USA, pp. 686–689, 2003.

Chapter 11

Dept. of Chemistry, Arizona State Univ., USA, available at http://www.chem.arizona.edu/massspec/intro_html/intro.html, on 30.10.2004.

Grattarola M., G. Massobrio, S. Martinoia, "Modeling H+-Sensitive FET's with SPICE", *IEEE T. on electron devices,* 39/4, pp. 813–819, 1992.

Research trends, A publication by Corporate Research, Special Edn., Oct. 2002.

Gutierrez-Osuna, R., "Machine Olfaction: Advanced excitation methods for inorganic chemoresistors", Wright State Univ.

Dutta, Ritaban, Evor L-Hines, Julian W-Gardner and Pascal Boilot, "Bacteria classification using Cyranose 320 electronic nose", Division of Electrical and Electronic Engineering, School of Engineering, Univ. of Warwick, United Kingdom, BioMedical Engineering OnLine 2002, 1:4.

Alegret, Salvador, "Chemical sensors as integrated analytical systems", Dept. de Química, Univ. Autonoma de Barcelona, 08193 Bellaterra Catalonia, Spain.

Trebotich, D., J, D. Zahn and D. Liepmann, "Complex Fluid Dynamics in BioMEMS: Modeling of Microfabricated Microneedles", *Modelling and Simulation of Microsystems,* 2002, ISBN 0-9708275-1.

Wróblewski, Wojciech and Zbigniew Brzózka, "Ion-selective field effect transistor for nitrite determination in flow-cell", Dept. of Analytical Chemistry, Warsaw Univ. of Tech., 00-664 Warszawa, Noakowskiego 3.

"Ion selective transistor modelling for behavioural simulations", Daniel M., M. Janicki, W. Wroblewski, A. Dybko, Z. Brzozka and A. Napieralski, Dept. of Microelectronics & Comp. Sc., Technical Univ. of Lodz, Al. Politechniki 11, 93–590 Lodz, Poland available at

http://www.sewing.mixdes.org/downloads/autmonet03/janicki_autmonet.pdf

http://www.microe.rit.edu/research/nanotechnology/dnasensor.htm

http://www.signal.uu.se/Courses/Descr9900/Sensors%2000/uS7-,bio,chemical.PDF

Hierlemann, Andreas, Oliver Brand, Christoph Hagleitner and Henry Baltes, Microfabrication Techniques for Chemical/Biosensors", *Proc. of the IEEE,* 91/6, June, 2003

Liu Juewen and Yi Lu, "A Colorimetric Lead Biosensor Using DNAzyme-Directed Assembly of Gold Nanoparticles", Dept. of Chemistry, Univ. of Illinois at UrbanasChampaign, Urbana, Illinois 61801, Received Feb. 20, 2003; E-mail: yi-lu@uiuc.edu, Published on Web 05/10/2003, 6642 9 *J. AM. CHEM. SOC. 2003,* 125, 6642-6643 10.1021/ja034775u CCC: $25.00 © 2003 American Chemical Society.

Thewes, R. et al., "Sensor Arrays for Fully Electronic DNA Detection on CMOS", *ISSCC Dig. of Tech. Papers,* 2002, p. 350.

Hintsche, R. et al, "Microbiosensors using electrodes made in Si-technology in Frontiers in Biosensorics I", F. Scheller et al., (Ed.), Birkhuser, Basel, Switzerland, 1997.

Schienle, M. et al., "A Fully Electronic DNA Sensor with 128 Positions and in-Pixel A/D Conversion", *ISSCC Dig. of Tech. Papers,* p. 220, 2004.

Bernstein Michael, m_bernstein@acs.org, 202-872-6042, American Chemical Society, "Tiny nanowire could be next big diagnostic tool for doctors", Public release date: 16-Dec-2003, http://www.eurekalert.org/pub_releases/2003-12/acs-nc121603.php

Judy, Jack W., "Biomedical applications of MEMS", Univ. of California, Los Angeles, Los Angeles, CA 90095-1594, jjudy@ucla.edu,

http://www.ee.ucla.edu/~jjudy/publications/Conference./msc_2000_judy.pdf

"DNA Testing: An Introduction for Non-Scientists", An Illustrated Explanation, by Riley, Donald E., Univ. of Washington, *Science testimony*.

Dickert, Franz L., Oliver Hayden and Konstantinos P. Halikias, "Synthetic receptors as sensor coatings for molecules and living cells", *J. of Royal Society of Chemistry,* p. 126, pp. 766–771, 2001

Hoummadyy, Moussa, Andrew Campitelliyz and Wojtek Wlodarskiyz, "Acoustic wave sensors: design, sensing mechanisms and applications, IOP Publishing Ltd, *Smart Mater. Struct,* 6, 1997, 647–657. 1997, Printed in the UK.

Wiley, W.C.; Mclaren, I.H.; *The Review of Scientific Instruments;* 26,12, 1955, pp1150–1157.

Cotter, R.J.; *Analytical Chemistry;* 64,21, 1992, pp. 1027A–1039A.

Mamyrin, B.A.; Karataev, V.I.; Shmikk, D.V.; Zagulin, V.A.; *Soviet Physics - JETP*; 37,1,, 1973, pp. 45–48 and Mamyrin, B.A.; *Int. J. of Mass Spectrometry and Ion Processes;* 131, 1994, pp. 1–19.

Datskos, Panos, "MEMS based Calorimetric Spectroscopy", Oak Ridge National Laboratory, http://www.mnl.ornl.gov

Crocombe, Richard A, "MEMS technology moves process spectroscopy into a new dimension", *Spectroscopy Euroe,* Axsun technologies, Inc, 1 Furtune Drive, USA.

Cole, Marina, Nicola Ulivieri, Jesús García-Guzmàn and Julian W. Gardner "Parametric model of a polymeric chemoresistor for use in smart sensor design and simulation", *Microelectronics J.,* 34/9, Sept. 2003, pp. 865–875.

Dahlin, A P, M Wetterhall, G Liljegren, S K. Bergström, P Andrén, L Nyholm, K E. Markides and J Bergquist, "Capillary electrophoresis coupled to mass spectrometry from a polymer modified poly, dimethylsiloxane, microchip with an integrated graphite electrospray tip", *Analyst*, p. 130, 2, pp. 193–199, 2005.

Chapter 12

http://www.e-paranoids.com/a/ah/aharonov_bohm_effect.html

Denko, K.K.Showa, Marinobu Endo, "Very Small Nano-Carbon-Composite Gear Developed", Kitagawa Industries Co. Ltd., Seiko Instruments Inc., Feb. 6, 2002.

Saito, R., M. Fujita, G. Dresselhaus, and M. S Dresselhaus, "Electronic structure of chiral graphene tubules", *Appl. Phys. Lett*. 60, 18, 4 May 1992, American Inst. of Physics.

Norrie, A. A., R. J. Ballagh and C. W. Gardiner, "Quantum Turbulence in Condensate Collisions: An Application of the Classical Field Method", *Physical Review Letters, Phys. Rev. Lett*. 94, 040401 Feb. 2005.

Seaburn, Kent, "nanotube.chirality", July 14, 2004, Dave Smith (Ed.), *Nanopedia; The encyclopedia of Nanotechnology,* available at
http://nanopedia.cwru.edu/PrintPage.php?page=nanotube.chirality

Hoenlein, W., F. Kreupl, G. S. Duesberg, A. P. Graham, M. Liebau, R. Seidel and E. Unger, "Carbon Nanotubes for Microelectronics: Status and Future Prospects", *Materials Science & Engineering*, C 23, 663, 2003.

Grace, Iain, "Investigation into the properties of carbon nanotubes", Joule Laboratory, Dept. of Physics, Univ. of Salford.

Bockrath, M., D.H. Cobden, P.L. McEuen, N.G. Chopra, A. Zettle, A. Thess, and R.E. Smalley, "Single-electron transport in ropes of carbon nanotubes", *Science 1997*, 275; 1922.

Tans, S.J., M.H. Devoret, H. Dai, A. Thess, R. Smalley, L.J. Geerlings, and C. Dekker, Individual single-wall carbon nanotubes as quantum wires, *Nature,* London, 386; 474, 1997.

Keren, Kinneret, Rotem S. Berman, Evgeny Buchstab, Uri Sivan, and Erez Braun, "DNA-Templated Carbon Nanotube Field-Effect Transistor", *Science,* Vol. 302, www.sciencemag.org, 21 Nov., 2003.

Liu, Xiaolei, Chenglung Lee, and Chongwu Zhou, "Carbon nanotube field-effect inverters", *Applied Physics Letters,* 79/20, Nov. 2001.

Rosenblatt, S., Y. Yaish, J. Park, J. Gore, V. Sazonova, P.L. McEuen, *Nano Lett.* 2; 869, 2002.

Javay, A., H. Kim, M. Brink, Q. Wang, A. Ural, J. Guo, P. McIntyre, P. McEuen, M. Lundstrom, H. Dai, Nat. Mater. 1; 241, 2002.

Ghani, T., S. Ahmed, P. Aminzadeh, J. Bielefeld, P. Charvat, C. Chu, M. Harper, P. Jacob, C. Jan, J. Kavalieros, C. Kenyon, R. Nagisetty, P. Packan, J. Sebastian, M. Taylor, J. Tsai, S. Tyagi, S. Yang, M. Bohr, *IEDM Tech. Dig.,* 415, 1999.

Chang, B. Yu, L., S. Ahmed, H. Wang, S. Bell, C. Yang, C. Tabery, C. Ho, Q. Xiang, T. King, J. Bokor, C. Hu, M. Lin, D. Kyser, *IEDM Tech. Dig.,* 251, 2002.

Doris, B., M. Ieong, T. Kanarsky, Y. Zhang, R.A. Roy, O. Dokumaci, Z. Ren, F. Jamin, L. Shi, W. Natzle, H. Huang, J. Mezzapelle, A. Mocuta, S. Womack, M. Gribelyuk, E.C. Jones, R. Miller, H.P. Wong, W. Haensch, *IEDM Tech. Dig.,* 267, 2002.

Yue, G.Z., Q. Qiu, B. Gao, Y. Cheng, J. Zhang, H. Shimoda, S. Chang, J.P. Lu, O. Zhou, "Generation of continuous and pulsed diagnostic imaging X-ray radiation using a carbon-nanotube-based field-emission cathode", *Appl. Phys. Lett.* 81, 2; 355, 2002.

Cheng, Y., O. Zhou, "Electron field emission from carbon nanotubes", *C. R. Physique 2003,* 4; 1021.

Baughman, R.H., C.X. Cui, A.A. Zakhidov, Z. Iqbal, G.M. Spinks, G.G. Wallace, A. Mazzoldi, D. De Rossi, A.G. Rinzler, O. Jaschinski, S. Roth, M. Kertesz, *Carbon nanotube actuator, Science*, 284; 1340, 1999.

Iyuke, S. E., A. B. Mohamad, A. A. H. Kadhum, W. R. W. Daud, R. Chebbi, "Improved membrane and electrode assemblies for proton exchange membrane fuel cells". *J. of Power Sources,* 114, 2; 195, 2003.

Nagy, P., R. Ehlich, L.P. Biró,_, J. Gyulai, "Y-branching of single walled carbon nanotubes", *Appl. Phys. A,* 70, 481–483, 2000.

Gojman, Benjamin, Happy Hsin, Joe Liang, Natalia Nezhdanova, Jasmin Saini, "Y-Junction Carbon Nanotube Implementation of Intramolecular Electronic NAND Gate", Aug. 13, 2004.

Grimm, D., R. B. Muniz, A. Latge, *Phys. Rev. B*, 68, art. No. 193407, 2003.

Lou, Z. S., Q. W. Chen, W. Wang et al., *Carbon*, 41, 3063, 2003.

Lee, C. J., J. *Park, Appl. Phys. Lett*., 77, 3397, 2000.

Neimark, A. V., S. Ruetsch, K. G. Kornev et al., *Nano letters*, 3, 419, 2003.

Ahlskog, M., E. Seynaeve et al., *Chem. Phys. Lett*., 300, 202, 1999.

Koshio, A., M. Yudasaka, S. Iijima, *Chem. Phys. Lett*., 356, 595, 2002.

Katayama, T., H. Araki, K. Yoshino, *J. Appl. Phys*., 91, 6675, 2002.

Wang, Y. Y., G. Y. Tang, F. M. Koeck et al., *Diamond and Related Materials*, 13, 1287, 2004.

Latgé, A., R. B. Muniz, and D. Grimm, "Carbon Nanotube Structures: Y-Junctions and Nanorings", *Brazilian J. of Physics*, vol. 34, no. 2B, 585, Brazil, June, 2004.

László P. Biró, Géza I. Márk, and Antal A. Koós, "Coiled carbon nanotube structures with supraunitary nonhexagonal to hexagonal ring ratio", *Physical Rev. B 66*, 165405, 2002.

Hod, Oded and Eran Rabani, "Carbon nanotube closed-ring structures", *Physics Review*, B 67, 195408, 2003.

Oudenaarden, A. van, M. H. Devoret, Yu. V. Nazarov, and J. E. Mooij, "Magneto-electric Aharonov-Bohm effect in metal rings", *Nature* 391, 768–770, 1998.

Aharonov, Y. and D. Bohm, "Significance of electromagnetic potentials in quantum theory", *Phys. Rev*. 115, 485–491, 1959.

Olariu, S. and I. Iovitzu Popèscu, "The quantum effects of electromagnetic fluxes", *Rev. Mod. Phys*. 57, 339–436, 1985.

London, F., "On the problem of the molecular theory of superconductivity", *Phys. Rev*. 74, 562–573, 1948.

Peshkin, M. and A. Tonomura, *The Aharonov-Bohm Effect*, Springer-Verlag: Berlin, 1989.

Peat, F. David, *Infinite Potential: The Life and Times of David Bohm*, Addison-Wesley: Reading, MA, ISBN 0-201-40635-7, 1997.

Sjöqvist, E., "Locality and topology in the molecular Aharonov-Bohm effect", *Phys. Rev. Lett*. 89, 21, 210401/1–3, 2002.

Chen, Qianwang, Zhao Huang, *Carbon-Based Nanostructures*, N.P. Mahalik (Ed.), Springer and Verlag, Germany.

Chapter 13

Mahadevan, Ramaswamy, NSF Sponsored Workshop on Structured Design Methods for MEMS Structured Design Methods for MEMS, MEMS Technology Applications Center, http://www.mcnc.org/HTML/ETD/MEMS/memshome.html

http://www.eng.buffalo.edu/%7Eabani/fem/node1.html

http://caswww.colorado.edu/courses.d/IFEM.d/Home.html
http://www.sv.vt.edu/classes/MSE2094_NoteBook/97ClassProj/num/widas/history.html
http://documents.wolfram.com/applications/structural/FiniteElementMethod.html
http://audilab.bmed.mcgill.ca/~funnell/AudiLab/teach/fem/fem002.html
amath.colorado.edu/faculty/ spraguma/MCEN4030/intro_lecture.pdf

Rensselaer Polytechnic Inst., Scientific Computation Research Center, available at, http://www.scorec.rpi.edu/research_multiscale.html

Lorenz, Gunar, Arthur Morris, Issam Lakkis, "A top-down design flow for MOEMS", Coventor, Inc.

Morris, Arthur, Stephen Bart, Derek Kane, Gunar Lorenz, Vladimir Robinovich, "A design flow for MEMS", Microcosm Technologies Inc.

Kim, Y. Y. and G. H. Yoon, "Multi-Resolution Multi-Scale Topology Optimization: A New Paradigm", *Int. J. Solids Structures*, Vol. 37, pp. 5529–5559, 2000.

Kim, T. S. and Y. Y. Kim, "Multiobjective Topology Optimization of a Beam under Torsion and Distortion",¡¡ *AIAA J.*, 40/2, pp. 376–381, 2002.

Kim, Yoon Young and Jang Gang-Won, ¡¡"Hat Interpolation Wavelet-Based Multi-Scale Method for Thin-Walled Box Beam Analysis", *Int. J. Num. Methods Eng.*, Vol. 53, 7, pp. 1575–1592, 2002.

Kim, Y. Y. and E.-H Kim, "Effectiveness of the Continuous Wavelet Transform in the Analysis of Some Dispersive Elastic Waves", *J. Acoust. Soc. Amer.*, Vol. 110, 1, pp. 86–94, 2001.

Kim, Y. Y. and Y. Kim, "A One-Dimensional Theory of Thin-Walled Curved Rectangular Box Beams under Torsion and Out-of-Plane Bending", *Int. J. Num. Methods Eng.*, Vol. 53, 7, pp. 1675–1693, 2002.

Chaturvedi, R., C. Huang, B. Kazmierczak, T. Schneider, J. A. Izaguirre, T. Glimm, H. G. E. Hentschel, S. A. Newman, J. A. Glazier5, M. Alber, "On Multiscale Approaches to Three-Dimensional Modeling of Morphogenesis": http://userwww.service.emory.edu/~tglimm/texte/3d_1.pdf

Weinan E, Princeton Univ., Bjorn Engquist, Univ. of Texas, Xiantao Li, Univ. of Minnesota, Weiqing Ren, Princeton Univ., Eric Vanden-Eijnden, New York Univ., "The Heterogeneous Multiscale Method: A Review".

Madureira, Alexandre L., Analysis, Numerical Methods, and Applications of Multiscale Problems", Laborat ́orio Nacional de Computac ̧ ̃ao Cient ́ ́zfica - LNCC, Brazil.

Asakuma, Yusuke, Munetaka Soejima, Tsuyoshi Yamamoto, Hideyuki Aoki and Takatoshi Miura, A New Estimation Method of Coke Strength by Numerical Multiscale Analysis", *ISIJ Int.*, 43/8, pp. 1151–1158, 2003.

Rudd, Robert E., "Coarse-Grained Molecular Dynamics and Multiscale Modeling of NEMS Resonators", Lawrence Livermore National Lab., 7000 East Ave, L-045, Livermore, CA 94550 USA.

Ghoniemy, Nasr M., Esteban P. Busso, Nicholas Kioussis, Hanchen Huang, "Multiscale modelling of nanomechanics and micromechanics: An overview", *Philosophical Magazine*, Vol. 83, Nos. 31–34, 3475–3528, Dec 2003.

Schmitz, T.L., J.A. Dagata and B. Dutterer, W. Gregory Sawyer, "A Multiscale Fabrication Approach to Microfluidic System Development", *J of Manufacturing Processes*, Vol. 6/No. 1, 2004.

Asthagiri, D. and D. Bashford, "Continuum and Atomistic Modeling of Ion Partitioning into a Peptide Nanotube", *Biophys J*, 82/3, pp. 1176–1189, March 2002.

Schulze, Tim P. a,, Peter Smereka b, Weinan E, "Coupling kinetic Monte-Carlo and continuum models with application to epitaxial growth", *J. of Computational Physics,* 189, 197–211, 2003.

Tarek, M., B. Maigret, and C. Chipot, "Molecular Dynamics Investigation of an Oriented Cyclic Peptide Nanotube in DMPC Bilayers", *Biophys. J.*, 85,4: 2287–2298. Oct. 1, 2003.

Liu, Wing Kam, and Harold S. Park, "Bridging Scale Methods for Computational Nanotechnology, *Handbook of Theoretical and Computational Nanotechnology* (*Eds*), Nov. 16, 2004.

Mortensen, J. J., J. Schiøtz and K. W. Jacobsen, SIMU Newsletter, issue 4, 2002.

Robert Calvet MEMS Simulation and Design Optimization with FEA-Based Tools, Optomechanical Engineer, SiWave, Inc., Arcadia, CA.

Clark, J. V., D. Garmire, M. Last, J. Demmel, "Practical Techniques for Measuring MEMS Properties", *Nanotech 2004*, The Technical Proc. of the Nano Science and Technology Inst., Boston, MA, Vol. 1, pp. 402–405, March 7-11, 2004.

Clark, J. V., D. Bindel, W. Kao, E. Zhu, A. Kuo, N. Zhou, J. Nie, J. Demmel, Z. Bai, S. Govindjee, K. S. J. Pister, M. Gu, and A. Agogino, "Addressing the Needs of Complex MEMS Design", MEMS, Las Vegas, Nevada, Jan. 20–24, 2002.

Clark, J. V., D. Bindel, N. Zhou, S. Bhave, Z. Bai, J. Demmel, and K. S. J. Pister, "Sugar: Advancements in a 3D Multi-Domain Simulation Package for MEMS", In: *Proc. of the Microscale Systems: Mechanics and Measurements*.

Clark, J. V., N. Zhou, K. S. J. Pister, "MEMS Simulation Using SUGAR v0.5, In: *Proc. Solid-State Sensors and Actuators Workshop*, Hilton Head Island, SC, June 8–11, pp. 191–196 1998.

Youn, Sung Kie, Kwak, Byung M., Kwon, Jang-Hyuk, Chang, Su-Young; Huh, Jae S.; Kim, Eugene, Efforts in developing design and simulation tools for MEMS: DS/MEMS and CA/MEMS, *Proc. SPIE*, Vol. 4755, p. 275–284, Design, Test, Integration, and Packaging of MEMS/MOEMS 2002.

Ortiz, M., A. M. Cuiti˜no, J. Knap_ and M. Koslowski, "Mixed Atomistic-Continuum Models of Material Behavior: The Art of Trascending Atomistics and Informing Continua".

MILLERa, RONALD E. and E. B. TADMORb, "The Quasicontinuum Method: Overview, applications and current Directions", *J. of Computer-Aided Materials Design*, 9: 203–239, Kluwer Academic Publishers. Printed in the Netherlands

Xiao, S.P., T. Belytschko, "A bridging domain method for coupling continua with molecular dynamics", *Comput. Methods Appl. Mech. Engrg*. 193, 2004.

Rudd, Robert E. and Jeremy Q. Broughton, "Coarse-grained molecular dynamics and the atomic limit of finite elements", *Physics Review B*, 58/10, Sept. 1998.

Park, Harold S., Eduard G. Karpov, Wing Kam Liu, "The Bridging Scale for Two-Dimensional Atomistic/ Continuum Coupling

Rudd, R. E. and J. Q. Broughtonb, "Coupling of Length Scales and Atomistic Simulation of MEMS Resonators", Dept. of Materials, Univ. of Oxford.

Arndt, Marcel, "Upscaling Technique for the Atomistic-Continuum Simulation of Shape Memory Alloys with the EAM Potential", *Computational Mechanics*, Tsinghua Univ. Press & Springer-Verlag, WCCM VI in conjunction with APCOM'04, Sept. 5–10, Beijing, China, 2004.

Rodney, D., "Mixed atomistic/continuum methods: Static and dynamic quasicontinuum methods", *Génie Physique et Mécanique des Matériaux*, France

Chapter 14

http://www.chinarel.com/LifeDataWeb/a_brief_introduction_to_reliability.htm

Zhang, Gang‡, Huikai Xie*, Lauren E. de Rosset* and Gary K. Fedder*†, "A lateral capacitive CMOS accelerometer with structural curl compensation," *Dept. of Electrical and Computer Engineering and †The Robotics Inst. Carnegie Mellon Univ., Pittsburgh, PA 15213–3890, ‡now with Motorola Inc., Semiconductor Products Sector, MD Z207, Phoenix, AZ 85008.

Mutlu, Senol and Brian Jensen, "Design and Simulation of Shaped Comb Fingers for Compensation of Mechanical Restoring Force", Univ. of Michigan.

MEMS Fatigue, MEMS Reliability Newsletter, A publication offered by Exponent. A World leader in MEMS reliability. 1/1, 21 Strathmore Road, Natick, MA 01760, Sept. 2001.

Nicholas, J. V.; D. R. White, *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*, 2nd Edn., Wiley Series in Measurement Science and Technology.

Jiantao Pan, "MEMS and Reliability", Carnegie Mellon Univ., Photon Technology Int.

Yang, Shao, Igor Vayshenker, Xiaoyu Li, and Thomas R. Scott, "Accurate measurement of optical Dei13cior nonmwmuty", *NCSL Workshop & Symp*., 353, 1994.

Fradkov, Alexander, "A Nonlinear Philosophy for Nonlinear Systems", Inst. for Problems of Mechanical Engineering, Russian Academy of Sciences.

Booton, R., Jr., "The Measurement and Representation of Nonlinear Systems", *IRE T. on Circuit Theory*, pp. 32–34, 1/4, Dec 1954.

Griffin, John Linwood, Steven W. Schlosser, Gregory R. Ganger, David F. Nagle, "Modeling and Performance of MEMS-Based Storage Devices", *Proc. of ACM SIGMETRICS 2000*, 28/1, pp 56–65, California, June 17–21, ISSN:0163-5999 2000.

Mustafa Uysal, Arif Merchant, Guillermo A. Alvarez, "Using MEMS-based storage in disk arrays", *Proc. of the 2nd USENIX Conf. on File and Storage Technologies*, San Francisco, CA, March 2003.

Pantazi, A. M.A. Lantz, G. Cherubini, H. Pozidis and E. Eleftheriou "A servomechanism for a micro-electromechanical-system-based scanning-probe data storage device", *Nanotechnology*, Inst. of Physics Publishing, 15, 2004.

Oak Ridge National Laboratory, "MEMS Bio-Chemical Transducer, Calorimetric MEMS sensor array platform", U. S. Patent No. 6,436,346, Aug., 2002.

Reed, Jeffery A., PhD thesis, "Frequency selective surface with multiple periodic elements", Univ. of Texas at Dallas, Dec. 1997.

Free-space optical switch, US patent No. 6,507,683, Jan. 14, 2003, Assignee, Mitsubishi Dencki Kabushiki Kaisha, Tokyo, JP, Inventors, Sugitatsu; Atsushi, Tokyo, JP, Saito; Takeshi, Tokyo, JP, Hatta; Tatsuo, Tokyo, JP.

Zendejas, Joe, John Gianvittorio, Bongyoung Yoo, Yahya Rahmat-Samii, Ken Nobe, and Jack W. Judy, "Ferromagnetic MEMS array for reconfigurable frequency selective surfaces", Univ. of California, Los Angeles, 420 Westwood Plaza, Los Angeles.

Smith J. H. et al., "Embedded Micromechanical Devices for the Monolithic Integration of MEMS with CMOS", *Dig. IEDM*, pp. 609–612. Dec. 1995.

Index

The McGraw-Hill Companies